ORIGINAL ARTICLE

Open Access

CrossMark

Modelling production-consumption flows of goods in Europe: the trade model within Transtools3

Gerard de Jong[1,2*], Reto Tanner[1,3], Jeppe Rich[4], Mikkel Thorhauge[4], Otto Anker Nielsen[4] and John Bates[5]

* Correspondence:
dejong@significance.nl
An earlier version of this paper was presented at the European Transport Conference 2016 (ETC 2016) in Barcelona.
[1]Institute for Transport Studies, University of Leeds, University Road 36-40, Leeds LS2 9JT, UK
[2]Significance, The Hague, The Netherlands
Full list of author information is available at the end of the article

Abstract

The paper presents a new model for trade flows in Europe that is integrated with a logistics model for transport chain choice through Logsum variables. Logsums measures accessibility across an entire multi-modal logistical chain, and are calculated from a logistics model that has been estimated on disaggregated micro data and then used as an input variable in the trade model. Using Logsums in a trade model is new in applied large-scale freight models, where previous models have simply relied on the distance (e. g. crow-fly) between zones. This linkage of accessibility to the trade model makes it possible to evaluate how changes in policies on transport costs and changes in multi-modal networks will influence trade patterns. As an example the paper presents outcomes for a European-wide truck tolling scenario, which showcases to which extent trade is influenced by such a policy. The paper discusses how such a complex model can be estimated and considers the choice of mathematical formulation and the link between the trade model and logistics model. In the outcomes for the tolling scenario we decompose the total effects into effects from the trade model and effects from the logistics model.

Introduction

Trade models can be used to forecast future trade patterns conditional on scenarios about the economic development of various regions. If they would contain transport time and cost as explanatory factors of the trade volumes, trade models could also be used to simulate the impact of changes in transport costs (e.g. introduction of road toll) or transport time (e.g. constructing new links or expanding existing links) on trade flows. However, most existing large-scale trade models use a simple (e.g. crow-fly) distance variable as the measure of resistance between zones on trade, not transport times and costs. Very little empirical material is available on the impact of changes in transport costs and times (by mode) on the trade flows, but the few studies that have been done show that these effects are potentially large (see de Jong et al. 2010). This paper presents a new European trade model that is integrated with a logistics model (and a network assignment model), where Logsums are consistently used at large-scale. Accessibility is measured in this model across an entire multi-modal logistical chain, on the basis of a logistics model which has been estimated on micro data. This makes it possible to evaluate how changes in policies and changes in multi-modal networks will influence trade patterns.

Springer Open

de Jong *et al. Journal of Shipping and Trade* (2017) 2:5

Page 2 of 23

The trade model presented in this paper is developed as part the European Transtools3 model. The Transtools3 model is a new forecasting model system for passenger and freight transport in Europe, developed for DG MOVE of the European Commission. It consists of three main blocks: a passenger transport model, a freight and logistics model and a network assignment model. This paper focuses on the trade model, a specific sub-model of the freight and logistics model, as depicted in the top-right box in Fig. 1. The trade model produces growth in goods flows between Production and Consumption zones (PC flows, measured in tonnes) between a base year and a future year. The freight model as a whole was based on the 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 the regional transport distribution of trade, e.g. PC flows. These are: i) Gravity models, ii) Input-output (IO) models and iii) 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 et al. 2008). In such models the flow between zone i and zone j is a function of production and attraction measures of zone i and zone j and some measure of distance or (generalised) transport cost in order to account for the fact that the transport distribution is sensitive to changes in transport costs. Gravity models, as a mean of explaining trade or transport in money or weight units, have long been regarded as models without a clear theoretical justification. However, 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. Whereas national input-output tables have been developed for many countries, a multi-regional or spatial input-output table does not usually exist. Such a table would not only include 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:
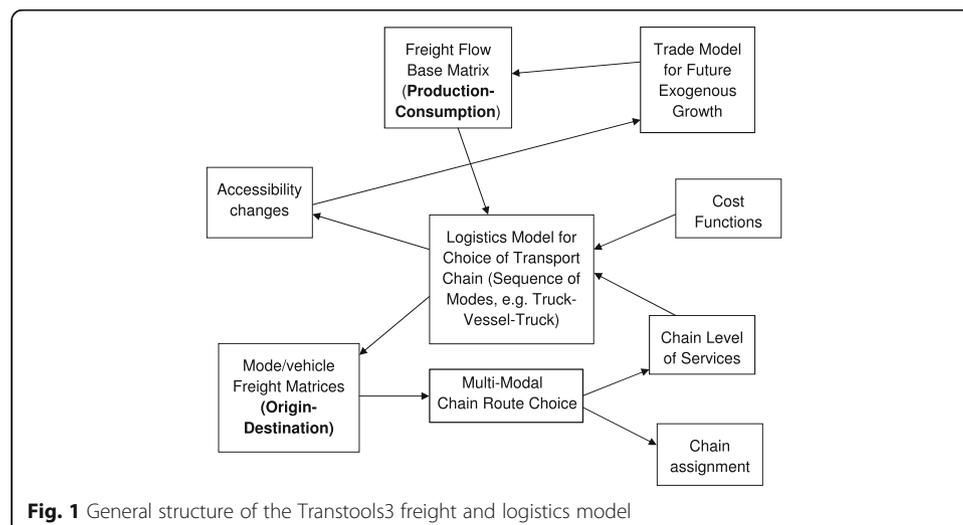

Fig. 1 General structure of the Transtools3 freight and logistics model

de Jong *et al. Journal of Shipping and Trade* (2017) 2:5

Page 3 of 23

– 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 productions and attractions is the computable general equilibrium (CGE) model that establishes equilibrium in several related markets (not only transport, but also goods markets, labour markets and possibly 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. Spatial computable general equilibrium (SCGE) models can be seen as a spatial extension of the CGE framework with at least the same data requirements as IO models (e.g. multiregional IO tables or make and use tables). Examples of operational SCGE 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 applied the gravity model instead of the IO and the SCGE models. The first reason for this choice is that in Transtools3, we had no ambitions to model markets other than transport. The second reason is that up-to-date information on regional input-output relations for our study area was missing for many regions. By contrast, as part of a previous data project (ETISplus 2014) a base PC matrix was made available. As this matrix was already expressed in tonnes, we avoided the conversion from money to tonnes, which would otherwise have been necessary prior to an IO and SCGE model application to transport.

The trade model in Transtools3 explains the transport flows by commodity type (NST/R 1) either between NUTS3 zones or between countries. It was estimated partly on data at the country level and partly at the NUTS3 level, and it is applied at the NUTS3 level. The reason for partly estimating on country level is discussed in chapter 3. The model is based on the (unconstrained) gravity formulation, using characteristics of zones and the transport costs between zones. The dependent variables are PC flows between zones (measured in tonnes). 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 effect of transport cost is integrated in a Logsum variable (see Ben-Akiva and Lerman 1985 for an exposition of the Logsum as an accessibility measure) which is calculated from the Transtools3 logistics model (see Jensen et al. A model for freight transport chain choice in Europe, submitted). It includes a range of different transport time components and transport costs components (for all modes considered).

The use of a Logsum variable as a measure of accessibility in a trade model and as a link between trade and logistics has been suggested in theoretical papers (e.g. de Jong and Ben-Akiva 2007). However, to our knowledge it has not yet been applied in large-scale practical freight transport models. To investigate how accessibility affects trade, we provide a European-wide model application of a truck toll scenario.

In the estimation of the trade model we also take into account the influence of relative trade resistance between countries instead of absolute resistance (in line with trade theory) by estimating a random effects model (see "Accounting for relative trade cost" subsection).

de Jong *et al. Journal of Shipping and Trade* (2017) 2:5

Page 4 of 23

The trade model is applied using a pivot point approach (Daly et al. 2011). This means that in the current model implementation we only use the trade model to predict relative changes over time that result from (scenario-based) changes in GDP by zone and in the population. These relative changes are then used to compute changes in the transport flows by commodity type and zone pair by applying these relative changes to the base PC matrix. All other variables are assumed to remain constant.

The key objective of this paper is to present the trade model component of the Transtools3 freight transport model and show how it was developed and which choices were made in the development process (and why). As a second objective, we want to show the sensitivity of this model to changes in policy-related variables.

In "The gravity model for trade" section of this paper, we present the gravity model for trade between zones. The data are described in "The data used" section. The logistics model section discusses key features of the logistics model. The estimation results for the trade model are presented in "Estimation results" section. The model implementation is described in "Model implementation" and "Conclusions" section offers the conclusions of the paper.

The gravity model for trade

Accounting for relative trade cost

We start with a trade model with distance as measure of resistance, at the country level.

The modern theoretical literature on gravity-based trade models (Anderson and van Wincoop 2003; 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 absorb 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:

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

(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_i: gross domestic product in euro of 2010;
pop_i: population;
$euefta_{ij}$: dummy that equals 1 if both countries are member of EU or EFTA; 0 otherwise;
$euro_{ij}$: dummy that equals 1 if both countries have the EURO as a currency; 0 otherwise;
$neig\_ij$: dummy that equals 1 if both countries are neighbours; 0 otherwise;
$lang\_ij$: dummy that equals 1 if both countries have the same language; 0 otherwise;

de Jong *et al. Journal of Shipping and Trade* (2017) 2:5

Page 5 of 23

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 that in this model we cannot estimate parameters for the destination zone variables gdp_j and $\frac{gdp_j}{pop_j}$ 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 thus reduce the variance of the error terms ε. Nonetheless, it 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.

Another way to take account of relative trade costs, which does not have this disadvantage, is the random effects model. The preferred trade model in Transtools3 was estimated as a random effects model for this reason. In this model, 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.

$$
\ln(x_{ij}) = \sum_{k=1}^{K_D} \alpha_k \ln(d_{ij(k)}) + \beta_1 \ln(gdp_i) + \beta_2 \ln(gdp_j) + \beta_3 \ln\left(\frac{gdp_i}{pop_i}\right) + \beta_4 \ln\left(\frac{gdp_j}{pop_j}\right)
$$
$$
+ \beta_5 euefta_{ij} + \beta_6 euro_{ij} + \beta_7 neig_{ij} + \beta_8 lang_{ij} + \vartheta_j + \varepsilon_{ij}
$$

$$(2)$$

Replacing distance by the Logsum variable

Equations (1) and (2) use a function of distance as the measure of resistance between zones i and j. Equation (3) below instead of distance includes the Logsum from the transport chain choice model.

$$
\ln(x_{ij}) = y_k LS_{ijk} + \beta_1 \ln(gdp_i) + \beta_2 \ln(gdp_j) + \beta_3 \ln\left(\frac{gdp_i}{pop_i}\right) + \beta_4 \ln\left(\frac{gdp_j}{pop_j}\right)
$$
$$
+ \beta_5 euefta_{ij} + \beta_6 euro_{ij} + \beta_7 neig_{ij} + \beta_8 lang_{ij} + \vartheta_j + \varepsilon_{ij}
$$

$$(3)$$

Where:

LS_{ijk}: Logsum variable for the accessibility of zone pair i-j for commodity type k. This variable is computed from the transport chain choice model that is also part of the Transtools3 freight transport model.

de Jong *et al. Journal of Shipping and Trade* (2017) 2:5

Page 6 of 23

Please note that in the estimation of the Logsum coefficients we use NUTS3 zones instead of countries, as the logistics model (which generates the Logsums) is estimated and applied at a zonal level.

The data used

Transtools3 uses a zoning system which is similar to the NUTS3 level, but sometimes, where NUTS3 zones are relatively large, contains subdivisions of NUTS3. Hence, the application of the trade model takes place at this rather detailed zoning level. This does however not require that all the estimation also uses this zoning level.

We received from the ETISplus project transport flows in tonnes, by NST/R 2 at the NUTS3 level ('zonal' level) for 2010. However, we prefer to use the country data for estimation, since at this level the flows are observed data, obtained from international organisations, and harmonised by ETISplus. To produce matrices at the zonal level, ETIS-plus 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 the production-consumption matrix (PC matrix) from ETISplus at the NUTS3 level (ETISplus 2014). Each observation covers the flow of a specific type of goods following the NST/R level 2 (NST/R 2) classification (which we aggregated to NST/R 1) from an origin zone to a destination zone. They used a PC matrix of observed data at country level and then, as noted, imputed trade flows for each individual pair of zones. We estimated models explaining this PC matrix (de Jong et al. 2016), as well as models explaining the matrix of flows aggregated to the country to country level (using 214 countries in total). As a check, we compared the results of estimation at the country level with those at the zonal level. Both gave elasticities of comparable magnitude.

Our main explanatory variables are GDP, GDP per capita and a measure of resistance to trade (in the model according to Eq. (2) we used country-country distance for this and in the model described in Eq. (3) we used the zone-zone Logsum from the transport chain choice model that is also part of the Transtools3 freight transport model). 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 (largely prepared manually, meaning that we coded these variables ourselves for all the zones), see list of variables following Eq. (1) in "The gravity model for trade" section.

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.

de Jong *et al. Journal of Shipping and Trade* (2017) 2:5

Page 7 of 23

The logistics model

The Logsum variables that are used as input in the trade model (in estimation and in application) come from the Transtools3 logistics model. The logistics model in Transtools3 represents a relatively advanced model in which the choice between multiple logistical chains is dealt with in a nested logit representation. The Transtools3 model is described in more detail in de Jong et al. (2016) and Jensen et al. (A model for freight transport chain choice in Europe, submitted) and extends previous regional and national models (Rich et al. 2011; de Jong and Ben-Akiva 2007; Ben-Akiva and de Jong 2013) to a wider geographic scope. It was estimated on data at the individual shipment level, gathered from shippers in France (the ECHO data) and Sweden (the 2009 Commodity Flow Survey). The dependent variable is the transport chain (a sequence of one or more modes) used between the production and the consumption location (consistent with the PC approach used in the trade model). Different models were estimated for:

1. Dry bulk
2. Liquid bulk
3. Containers and general cargo.

The choice alternatives in the logistics model are:

- Road direct (includes road-ferry combinations) – container
- Road direct (includes road-ferry combinations) – non-container
- Road with roll on/roll off (RORO) – container
- Road with RORO – non-container
- Rail – container
- Rail – non-container
- Inland waterways (IWW)
- Rail and IWW
- Sea
- Rail and sea
- IWW and sea
- Rail and IWW and sea.

Road can be included in all these alternatives (e.g. as access/egress mode). It is assumed that non-bulk goods transport by sea or IWW will be in containers. For road and rail there is the choice in the model between transporting this as general cargo or in containers.

Finally, the influencing factors in the logistics model are:

- Transport cost
- Transport time
- Commodity type
- Value density of the goods
- Direct access to rail or IWW.

The Logsum from the logistics model provides a measure of accessibility over all available transport chains between any two zones: it is the expected maximum utility

de Jong *et al. Journal of Shipping and Trade* (2017) 2:5

Page 8 of 23

from transport chain choice. Since the logistics model is a logit model, the Logsum is the denominator of the choice probability, which includes a summation over all transport chain alternatives. The Logsum will respond to changes in the above influencing factors of the logistics model.

Estimation results

During the estimation of the trade-model it has been considered whether we should estimate the model at the level of the NUTS3 zones or at the level of the countries. We 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, since the country level reflects the level at 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 effects model at country level and then fixed these in a subsequent estimation at the zone level (the second stage). The resistance variables in this first stage are distance splines, which is a much more flexible specification than its competitors (e.g. linear, logarithmic, quadratic) and explains the data clearly better than using continuous distance. In the second-stage estimation, we apply the parameters from the first stage as regards 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 would virtually "destroy" the variation in the Logsum variables from the logistics model (they vary a lot within countries).

Estimation stage 1

All models are estimated per NST/R 1 commodity type (10 models). 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 (non-zero 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 showed that the origin GDP elasticities do not change much compared to the models without fixed effects. Replacing GDP by (sectoral) gross value added did not improve the fit of the models or the significance of the estimates.

The estimation results of the random effects model with distance splines are given 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). After 2000 km (for many commodities) the impact of additional distance levels off. 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

Table 1 Estimation results for a model at the country level with country-specific random effects at the destination and distance as splines

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 0–20 km	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 20–50 km	0	0	0	0	0	0	0	0	0	0
	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)
Distance 50–100 km	0	0	0	0	0	0	0	0	0	0
	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)
Distance 100–300 km	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 300–500 km	−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 500–1000 km	−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 1000–2000 km	−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*

Table 1 Estimation results for a model at the country level with country-specific random effects at the destination and distance as splines (*Continued*)

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
	(−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
+ $p < 0.10$, * $p < 0.05$

de Jong *et al. Journal of Shipping and Trade* (2017) 2:5

Page 11 of 23

therefore derive the total GDP effect on trade from the model (while still taking account of destination-specific effects). In the random effects model we do not have to exclude destination (or origin) GDPs, in contrast to the fixed effects model where this is necessary because of the correlation between destination-specific fixed effects and GDP per destination (see "Accounting for relative trade cost" subsection). The total elasticities of a rise in GDP at fixed population (sum of all GDP and GDP/capita elasticities) are shown in Table 2. The variation by type of commodity is quite plausible. Note that the value for fertilisers is negative, 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.

Besides the random effects model, we also estimated a Heckman model (Heckman 1979), that explicitly takes account of the fact that many relations have no trade in a product. This model contains two related choices, one discrete choice to participate in trade ("selection") and one continuous choice on the amount of trade (when positive; "demand"). In the Heckman model some of the GDP elasticities of the demand equation (see Appendix 1) are higher now than in the random effect model, but many are also rather similar to the model in Table 1 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 smaller. It is not possible to have random effects and the Heckman specification at the same time (in the software used: Stata). 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. Given this finding, and the finding that most elasticities were rather similar, we decided to implement the random effects model in Transtools3.

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

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

de Jong *et al. Journal of Shipping and Trade* (2017) 2:5

Page 12 of 23

Estimation stage 2

The estimated Logsum parameters for stage 2 appear in Table 3 below. All the parameters are significant and with the correct sign, with the exception of the parameter for NSTR2 (Solid mineral fuels) which was found to be negative and not significant. This is not surprising given the earlier results and it has been decided to fix this value to 0. The model estimated in stage 2 also has an additional dummy variable for trade within countries.

Model implementation

Implementation method and elasticities

In the application of the trade model, apart from the Logsums, we only use the GDP and GDP per capita elasticities from Table 3, assuming that the dummies do not change (although in principle the model can also be used 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 applies the 2010 base PC matrix and income and population changes per zone. The base PC matrix, which was developed in the ETISplus project, includes the variables as described below in Table 4.

Within each scenario run with Transtools3, we need as model inputs 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):

$$\%\Delta gdp_P = \%\Delta gdp_C$$
$$\%\Delta pop_P = \%\Delta pop_C \qquad (4)$$

From these, we can also calculate the percentage growth in GDP per capita. In addition, a change in the Logsum between a production and consumption zone influences the trade level between the zones. In order to account for such an effect, the percentage change in Logsums between a production and consumption zone is calculated.

The trade model explains the percentage change (due to income and population change per zone and Logsum change) for each PC matrix cell value F_{PCg}. This relative change is then applied together with the base PC matrix to obtain the future year PC matrix. For each NST/R commodity class g we have:

$$\%\Delta F_{PC_g} = E_g(gdpcap_P)\%\Delta gdpcap_p + E_g(gdpcap_C)\%\Delta gdpcap_C + E_g\left(logsum_{CP_g}\right)\%\Delta logsum_{CP_g} \qquad (5)$$

Where $E_g(y_z)$ above denotes the estimated elasticity for changes in y (e.g. GDP per capita) 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 taken directly from the random effects model (presented in Table 1) as it is a log-log representation where elasticities are essentially identical to parameters. They have been estimated as independent GDP effects of P and C zones, and can therefore be added to give the total effect (as in Eq. (5)), even though in reality the underlying GDP trends of regions can be correlated.

The elasticities for the Logsums on the other hand, cannot be transferred directly (because we are interested in the impact of changing the Logsum as such by some percentage, not in changing the sum of which the logarithm is taken in the Logsum). Thus, the Logsum elasticities are computed using simulation. More specifically, we compute an increase in LN(tonnes) between zone pairs due to a 10% increase of the

Table 3 Trade model with random effects with estimation results for Logsums at the NUTS3 level (results of step 2 of the 2-step procedure)

	(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
both EU or EFTA	0.743	1.103	0.264	−1.204	0.357	0.435	0.357	−0.062	0.215	1.183
both EURO	0.596	−0.015	0.245	−0.298	0.562	0.124	0.423	1.191	0.324	0.181
neighbour countries	1.734	1.287	0.949	1.933	1.57	1.184	1.72	1.73	1.037	0.659
same language	0.742	1.02	0.899	0.968	0.903	0.671	0.777	0.106	0.945	0.963
same country	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			

de Jong *et al. Journal of Shipping and Trade* (2017) 2:5

Page 14 of 23

Table 3 Trade model with random effects with estimation results for Logsums at the NUTS3 level (results of step 2 of the 2-step procedure) *(Continued)*

	(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
LogSum_NSTR_7							(373.995)	0.397 (261.085)		
LogSum_NSTR_8									0.41 (470.569)	
LogSum_NSTR_9										0.29 (327.156)
Intercept	−9.538 (−105.409)	−9.504 (−118.787)	−14.393 (−126.565)	−12.59 (−185.125)	−11.417 (−64.125)	−12.199 (−135.462)	−10.299 (−94.105)	−0.816 (−8.148)	−8.78 (−83.572)	−12.604 (−161.173)
Observations	1,144,042	1,281,877	356,789	1,249,786	636,420	1,005,574	1,150,606	491,506	1,197,438	1,472,537
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

de Jong *et al. Journal of Shipping and Trade* (2017) 2:5

Page 15 of 23

Table 4 Data structure of trade matrices

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	

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 for the different variables is computed as:

$$E\left(logsum_{CP_g}\right) = \left(100 \frac{logsum_{CP_g} - logsum0_{CP_g}}{logsum0_{CP_g}}\right)/10 \tag{6}$$

Where in (6) $logsum0_{CP_g}$ represent the baseline Logsum and $logsum_{CP_g}$ the corresponding scenario variable. The overall elasticity measure for each commodity type g is computed based on a weighted average with respect to the amount of freight between zone pairs. The elasticities for GDP, GDPCAP, and Logsums are listed in Table 5.

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 of the underlying LoS variables, which can be interpreted.

The future flow *F(new)* can be calculated as:

$$F_{PC_g}(new) = F_{PC_g}(base)\left(100 + \%\Delta F_{PC_g}\right)/100 \tag{7}$$

Where $\%\Delta F_{PC_g}$ represent the percentage change in flow for commodity g and $F_{PCg}(base)$ is the base flow.

The output of the trade model consists of a new PC matrix (for a scenario in a future year).

The trade model is then followed by the logistics model (see Fig. 1 and "The logistics model" section). We considered doing the application of the transport chain model by means of 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 NST/R 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).

Table 5 Elasticities for the trade model

	NSTR 0	NSTR 1	NSTR 2	NSTR 3	NSTR 4	NSTR 5	NSTR 6	NSTR 7	NSTR 8	NSTR 9
$E\left(gdp_{P_g}\right)$	0.824	0.899	0.474	0.735	0.587	0.921	1.054	−0.002	1.123	1.175
$E\left(gdp_{C_g}\right)$	0.598	0.618	0.430	0.499	0.532	0.781	0.625	0.619	0.931	0.814
$E\left(gdpcap_{P_g}\right)$	−0.274	−0.211	−1.051	−0.425	−0.352	−0.393	−0.556	−0.627	0.119	−0.018
$E\left(gdpcap_{C_g}\right)$	−0.111	−0.094	0.422	0.316	−0.039	−0.094	−0.163	−0.390	−0.137	−0.104
$E\left(Logsum_{CP_g}\right)$	0.533	0.141	0.000	0.079	0.110	0.740	0.209	0.322	0.174	0.164

de Jong *et al. Journal of Shipping and Trade* (2017) 2:5

Page 16 of 23

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.

A European-wide kilometre-based charging experiment for road transport

To further test the trade model we have carried out a European-wide kilometre-based truck toll experiment within the Transtools3 freight model. A simulation of the impact of a toll on trade would not be possible in a trade model with distance as the only resistance term. Of course it is possible to increase the distances (for the zone pairs with a toll) as a proxy, but not all transport costs are distance-dependent and it would be unclear by how much the distances should be increased to mimic the toll. The Logsum on the other hand makes it possible to simulate both changes in time and in costs (or more generally; any attribute that is includes in the utility function of the logistics model), since the transport chain choice model includes both of these factors separately.

Firstly, we have constructed a reference case, which reflects the base case (2010) tolling levels in Europe as shown below in Table 6. The tolls in Table 6 are based on existing tolls, however for countries with hourly truck tolls these have been converted to a kilometre-based toll by assuming 80 km/h. Clearly, this is likely to be an overestimate as it neglects resting hour restrictions; however, it is reasonable as a means to test the sensitivity of the model to tolling.

For countries not listed in the reference scenario in Table 6 we assume a toll cost of 0. The two additional scenarios are then as follows. In the first, scenario A, we assume

Table 6 Reference scenario for kilometre-based truck tolls, source: Hylen et al. (2013)

Country	Type of tolling	Baseline (Euro/KM)
Austria	KM based	0.35
Czech Republic	KM based	0.26
France	KM based	0.2
Germany	KM based	0.18
Greece	KM based	0.16
Italy	KM based	0.13
Poland	KM based	0.09
Portugal	KM based	0.09
Russia	KM based	0.09
Slovakia	KM based	0.19
Slovenia	KM based	0.22
Spain	KM based	0.17
Switzerland	KM based	0.61
Belgium	Hourly based	0.006
Denmark	Hourly based	0.006
Hungary	Hourly based	0.00375
Lithuania	Hourly based	0.00375
Luxembourg	Hourly based	0.006
Netherlands	Hourly based	0.006
Sweden	Hourly based	0.006

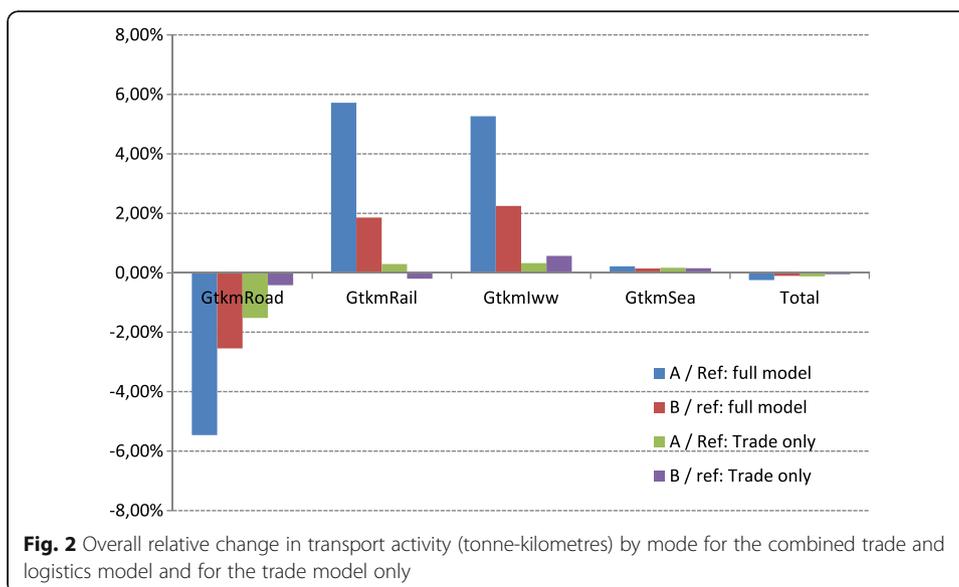de Jong *et al. Journal of Shipping and Trade* (2017) 2:5

Page 17 of 23

Table 7 Results for kilometre-based truck tolls by transport modes and model

Comparison	GtkmRoad	GtkmRail	GtkmIww	GtkmSea	Total
A / Ref: full model	−5.46%	5.72%	5.26%	0.22%	−0.25%
B / ref.: full model	−2.55%	1.85%	2.25%	0.14%	−0.10%
A / Ref: Logistic only	−3.93%	5.42%	4.94%	0.05%	−0.12%
B / ref.: Logistic only	−2.13%	2.05%	1.68%	−0.01%	−0.05%
A / Ref: Trade only	−1.52%	0.30%	0.33%	0.17%	−0.12%
B / ref.: Trade only	−0.42%	−0.20%	0.57%	0.15%	−0.05%

that the Austrian toll level applies to all of Europe, whereas, in the second, scenario B, we assume that the German toll level applies to all of Europe. These tolling levels also apply to countries which in the reference case had a charge of 0.

Hence, according to these scenarios, scenario A represents a relative aggressive tolling scenario where all other countries except Switzerland experience a cost increase. Scenario B also represents a cost increase on average although slightly less aggressive. However, the scenario is interesting as Germany is by far the biggest country in Europe when it comes to road transport. Other countries often act as price-followers when defining their respective tolling schemes.

The overall results are presented in Table 7 and Fig. 2 below. We present the model results for the combined effects of a logistic model and the trade model, as well as the isolated effects for the trade model. As can be seen, the road transport activity for the combined model is reduced by 5.5% for the aggressive scenario A and around half of that for scenario B. For road the trade model is responsible for around 20% of this effect, and for significantly less in the case of rail. For sea however, the pure trade effects are generally substantially larger. This is logical as most chains that go by sea have very limited logistical alternatives and this leads to a relative high impact of trade. The IWW alternative actually has a higher share for scenario B compared to A, which at first might be considered counterintuitive. However, the toll as described above has a very heterogeneous geographical nature which, combined with the fact that IWW is itself a very heterogeneous

Fig. 2 Overall relative change in transport activity (tonne-kilometres) by mode for the combined trade and logistics model and for the trade model only

de Jong *et al. Journal of Shipping and Trade* (2017) 2:5

Page 18 of 23

transport mode in its geographical scope, may produce such effects. The effect is also driven by the fact that IWW is often used in combination with other modes.

Clearly, the "trade-only" model is somewhat artificial in that we disentangle trade effects at the level of modes (whereas Logsums aggregate over the modes to provide a summary measure over all modes). This is achieved by simply running the model with all Logsum parameters at zero, thus bypassing level-of-service effects in the trade model to give the logistics model only effects at the mode level. We can then compare this with a model where Logsum variables are active to yield the trade model only effects. This makes it possible to present the results by activity.

Conclusions

This paper has presented a new European trade model that is integrated with a logistics model, where Logsums are consistently used at large-scale. Accessibility is measured in this model across an entire multi-modal logistical chain, on the basis of a logistics model which has been estimated on micro data. This makes it possible to evaluate how changes in policies and changes in multi-modal networks will influence trade patterns. Most existing large-scale trade models use a simple (e.g. crow-fly) distance variable as the measure of resistance between zones. This makes it hard to simulate the impact of changes in transport costs and transport networks on the trade flows.

The paper has discussed the existing literature on gravity-based trade models. It described the data and model structures used and presented the estimation results for random effects specifications with a either distance splines or Logsums as the measure of resistance to trade. Overall 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 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. Due to this the Transtools3 model has applied a random effects model.

The estimation of the trade model involved a number of considerations in terms of the level of estimation. It was decided to apply a two-stage approach, where in the first stage we estimated a generic random-effects model at the level of the countries. As the trade data originates at the country-to-country level it is natural to estimate GDP and country specific border effects at this level. These parameters were then transferred to a second stage estimation where we estimated regional variables and accessibility effects through a Logsum variable.

In Transtools3 the random effects model is used in the implementation of the freight and logistics model. Through the Logsum variable from the logistics model, there is an influence of transport cost and time on the pattern of PC flows, and not only on the choice of transport chain for each given PC flow.

As a final assessment of model sensitivity, we analysed two truck toll scenarios against a reference tolling scenario. For these three scenarios we compared a complete model run involving both logistics and trade effects with a model run where only the trade model was allowed to change. This allowed us to disentangle the isolated effects from the trade model in the final model framework. Results indicate that logistic effects are dominating although trade effects are substantial.

de Jong *et al. Journal of Shipping and Trade* (2017) 2:5

Page 19 of 23

Appendix 1

Accounting for relations with zero trade

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 manufacturer 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:

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

(8)

The selection equation is:

$$s_{ij} = \gamma_1 \ln\left(d_{ij}\right) + \delta_1 \ln\left(gdp_i\right) + \delta_2 \ln\left(gdp_j\right) + \delta_3 \ln\left(\frac{gdp_i}{pop_i}\right) + \delta_4 \ln\left(\frac{gdp_j}{pop_j}\right) + \varepsilon_{ij}$$

(9)

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. (2) we include the dummy for the same language for this, as an example. In the estimation in "The logistics model" section, we use all the dummies from Eq. (2) also in Eq. (9), but not in Eq. (8).

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.

Table 8 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 – 1000 km	−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 – 2000 km	−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)
> 2000 km	−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 – 1000 km	−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 – 2000 km	−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)
> 2000 km	−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)

Table 8 Estimation results based on the Heckman-model at the country level (*Continued*)

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
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 EU or EFTA	0.832*	0.725*	0.0739	0.595*	0.325*	0.611*	0.595*	0.202*	1.359*	7.693
	(6.95)	(5.48)	(0.96)	(6.34)	(4.06)	(5.16)	(5.89)	(2.68)	(6.10)	(0.00)
Both EURO as currency	0.199	0.201	0.589*	−0.295*	0.348*	0.198	0.0554	0.125	0.280	0.624
	(1.06)	(0.97)	(6.07)	(−2.30)	(3.06)	(1.11)	(0.38)	(1.20)	(0.76)	(0.00)
Neighbour countries	0.430	0.638	0.861*	0.617*	0.231	0.000327	0.674+	0.334*	−0.173	6.109
	(0.99)	(1.45)	(5.78)	(1.97)	(1.30)	(0.00)	(1.89)	(2.04)	(−0.48)	(0.00)
Both same language	0.299*	0.268*	0.180*	0.295*	0.263*	0.155+	0.388*	0.367*	0.458*	0.153+
	(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	10,205	10,205	10,205	10,205	10,205	10,205	10,205	10,205	10,205	10,205

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

de Jong *et al. Journal of Shipping and Trade* (2017) 2:5

Page 22 of 23

Authors' contributions
GdJ supervised the research project and took the lead in writing the paper. RT estimated all models except the one with the logsums. JR and MT estimated the model with logsums and performed the policy simulation on tolling. OAN is the Transools3 overall project leader and JB provide advice on the model specification. All authors read and approved the final manuscript.

Ethics approval and consent to participate
Not required.

Consent for publication
Consent for publication is granted and all data and materials used can be made available.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]Institute for Transport Studies, University of Leeds, University Road 36-40, Leeds LS2 9JT, UK. [2]Significance, The Hague, The Netherlands. [3]Delft University of Technology, Delft, The Netherlands. [4]DTU Technical University of Denmark, Kongens Lyngby, Denmark. [5]John Bates Services, Abingdon, UK.

References
Anderson JE, van Wincoop E (2003) Gravity with gravitas: a solution to the border puzzle. Am Econ Rev 93-3:171–192
Ben-Akiva ME, de Jong GC (2013) The aggregate-disaggregate-aggregate (ADA) freight model system. In: Ben-Akiva ME, Meersman H, van de Voorde E (eds) Freight transport Modelling. Emerald, Bingley
Ben-Akiva ME, Lerman SR (1985) Qualitative choice analysis: theory and application to travel demand. MIT Press, Cambridge
Bergstrand JH, Egger P, Larch M (2013) Gravity redux: estimation of gravity equation coefficients, elasticities of substitution, and general equilibrium comparative statics under asymetric bilateral trades costs. J Int Econ 89:110–121
Bröcker J, Korzhenevych A, Schuermann C (2010) Assessing spatial equity and efficiency impacts of transport infrastructure projects. Transp Res B 44:795–811
Daly AJ, Fox J and Patruni B (2011) Pivoting in travel demand models. Paper presented at the European Transport Conference 2011, Glasgow
de Jong GC, Nielsen OA, Henriques M, Tanner R, Johnson D, Sowa V, Vierth I, Bates J (2016) Deliverable 7.2 - Final report on the freight and logistics model. Seventh Framework Programme theme [SST.2010.1.3–1.], MOVE/FP7/266182/TRANSTOOLS 3, DTU Transport, Copenhagen. http://www.transportmodel.eu/system/files/TT3%20WP7%20_D7.2_freight%20and%20logistics%20model_3.0_Final.pdf. Accessed 18 Oct 2017.
de Jong GC, Schroten A, van Essen H, Otten M, Bucci P (2010) The price sensitivity of road freight transport – a review of elasticities. In: E. van de Voorde, TH. Vanelslander (eds) Applied transport economics, a management and policy perspective. De Boeck, Antwerpen
Deardorff A (1998) Determinants of bilateral trade: does gravity work in a neoclassic world? In: Frankel JA (ed) The regionalization of the world economy. University of Chicago Press, Chicago
Eaton J, Kortum S (2001) Trade in capital goods. Eur Econ Rev 45:1195–1235
Eaton J, Kortum S (2002) Technology, geography, and trade. Econometrica 70:1741–1779
Edwards H with assistance from Bates J and Swahn H (2008) Swedish base matrices report – estimates for 2004, methodology, data and procedures. Stockholm, SIKA PM
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
Gomez Herrera E (2010) Comparing alternative methods to estimate gravity models of bilateral trade. Department on Economic Theory, University of Granada, Granada
Heckman JJ (1979) Sample selection error as a specification bias. Econometrica 47:153–161
Hylen B, Kauppila J, Edouard C (2013) Road Haulage Charges and Taxes, Summary analysis and data tables 1998–2013, OECD, discussion paper 2013.08
Ivanova O (2014) Modelling inter-regional freight demand with input-output, gravity and SCGE methodologies. In: Tavasszy LA, de Jong GC (eds) Modelling freight transport, Elsevier Insights Series. Elsevier, Amsterdam
Ivanova O, Heyndrickx C, Spitaels K, Tavasszy L, Manshanden W, Snelder M, Koops O (2007) RAEM version 3.0. Transport Mobility Leuven, Leuven
Ivanova O, Vold A, Jean-Hansen V (2006) PINGO – a model for prediction of regional and interregional freight transport, TØI report 578/2002 (revised). Institute of Transport Economics, Oslo

de Jong *et al. Journal of Shipping and Trade* (2017) 2:5

Page 23 of 23

Jong GC d, Ben-Akiva ME (2007) A micro-simulation model of shipment size and transport chain choice. Transp Res B 41:950–965

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

Linders G-JM, de Groot HLF (2006) Estimation of the gravity equation in the presence of zero flow, Tinbergen Institute Discussion Paper TI 2006–072/3. VU University Amsterdam, Amsterdam

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

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

Rich J, Kveiborg O, Hansen CO (2011) On structural inelasticity of modal substitution in freight transport. J Transp Geogr 19(1):134–146

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