Summary

Norwegian Logistics Model: Moving from a deterministic framework to a random utility model

TØI Report 1538/2016
Authors: Elise Caspersen, Bjørn Gjerde Johansen, Inger Beate Hovi, Gerard de Jong
Oslo 2016 43 pages English language

In the present report we describe work related to estimating an econometric model for the choice of transport mode and shipment size, and how this econometric model can be used to improve the Norwegian Logistics Model. We have estimated discrete choice models for three different commodity groups using data from the 2009 Swedish Commodity Flow Survey. The estimated coefficients and their corresponding selection criteria functions have been implemented in the Norwegian Logistics Model by creating a new version of the executable ChainChoice program. We validated the model by comparing the deterministic model to the stochastic model through the demand elasticities for the three transport modes road, sea and rail, with respect to changes in time and distance-based link costs. The comparison showed that all own elasticities and all cross elasticities for sea and rail have the expected sign. For two of the commodity groups, the own elasticities in the stochastic case are consistently lower than the own elasticities in the deterministic case, while there is no clear pattern for the third commodity group.

Introduction

The purpose of this report is twofold. First, we describe the work related to estimating an econometric model for the choice of transport mode and shipment size in Norway, and then how this econometric model can be used to improve the Norwegian Logistics Model. The models presented in this report are estimated for three of the commodity groups in the current Norwegian Logistics Model. These are commodity group 13 “Iron and steel”, 17 “Plastic and rubber” and 30 “Consumables”. The models are estimated on shipment level data from the 2009 Swedish commodity flow survey (CFS).

When the logistics model for Norway was first conceived, the idea was to estimate a model based on data from the Swedish Commodity Flow Survey and a couple of logistics providers in Norway. However, since a deterministic logistics module is complex and estimation of disaggregate models take a significant amount of time, a ‘preliminary’ or ‘prototype’ version of the logistics model was developed (see de Jong and Ben-Akiva, 2007, section 8) in 2005-2006. Despite several improvements, the model has remained deterministic, relying on a cost minimisation procedure per firm-to-firm (f2f) to find preferred transport mode and shipment size. Recently, studies have shown that this assumptions is not always valid. Hence, it seems to be a sensible time to move from a deterministic framework to a random utility model for the Norwegian Logistics Model.

Econometric specification of a stochastic logistics model

One of the main reasons to move away from a deterministic model is that a deterministic model has weak empirical foundation in observed behaviour. In addition, given a deterministic cost-minimizing model it is difficult to get full information about all cost elements and other factors transport agents consider when making their choice. Also, if the relevant part of the logistics costs function is rather flat, only a small change in logistics costs can result in a shift to a completely different optimal shipment size and transport chain. Some of these issues can be solved by estimating disaggregate random utility models with available RP data. By their nature, such models are probabilistic models, because they include a stochastic component to account for the influence of omitted factors. We start the process of moving to a random utility model by estimating a joint decision model for three of the commodities in the Norwegian Logistics Model. Every commodity type for which we can base the choice mechanism on observed RP data, constitutes an improvement relative to the deterministic model.

The chosen estimation approach is a joint model with discrete mode and discrete shipment size choice. The advantage with this model specification is that it does not require a combination of techniques from discrete choice and regression analysis, but can be estimated fully within the discrete choice framework. This is also the chosen estimation technique in most of the existing studies in the field. The drawback is that if the choice of shipment size in reality is a continuous one, the approach may lead to measurement errors. In this case, an alternative model specification is a model with discrete transport mode and continuous shipment size. For our chosen model, one might overcome this problem by estimating a nested or a cross-nested logit model.

Data analysis and model approach

The dataset used for estimations is shipments between Norway and Sweden registered in the 2009 Swedish CFS. In total the data set covers 105,533 shipments between Norway and Sweden. For each shipment, the data contains information regarding size, value, mode of transportation, commodity group, and geographical location of senders and receivers of goods, with municipality as the lowest geographical level. Information about transport cost is missing, and needs to be calculated. This is done for both each observed shipment mode choice and its alternatives using the Norwegian Logistics Model. To estimate cost data, we need information about sending and receiving zones in both Norway and Sweden. Observations lacking information about one or both of these variables are excluded from the dataset.

The choice set used in the estimation is determined by all chosen combinations of transport chains and shipment size categories in the CFS, under the restriction that cost data can be estimated. For commodity group 13 Iron and steel 44 % of the shipments are transported by road, and 55 % by road-rail-road. Majority of the shipments weight more than 50 ton. For commodity group 17 Plastic and rubber 61 % is transported by road, while 38 % is transported by rail. Less than 1 % is transported by air or waterborne transport. Commodity group 17 is transported in smaller shipments than iron and steel. The choice set for commodity 30 Consumables consists of 12 available chain choice alternatives. This is the only (out of three) commodity group for which we have a sufficient number of observations to include water as a transport mode choice (in combination with road/rail).

Exogenous variables are transport cost, time use, degradation and capital cost, value density, inventory in transit and region specific dummies.

Discrete choice model estimation for the joint decision of transport mode and shipment size

Estimations are carried out for the three commodity groups using the 2009 Swedish Commodity Flow Survey, the defined choice set and a set of explanatory variables. All models are estimated with the software Biogeme (Bierlaire, 2003). In addition to the three multinomial logit models, we extend the estimation of commodity group 17 to a model where the utility functions are non-linear in cost and time. This extended model is estimated as both a multinomial logit and a nested logit model. This was not done for the other two commodity groups. Commodity 13 lacks observations for the alternative with small shipments transported by rail, while commodity 30 contains so many observations that extending the model led to a severe slowdown of the estimation in Biogeme.

We were not able to estimate a reliable nested logit model with non-linearity in the utility function. Since the multinomial logit model is a restricted version of the nested logit model, this also gives reasons to doubt the estimation of the multinomial model with non-linear variables. For this reason, we chose to keep the simpler models for the discrete choice, where time and cost are linear in the utility function, instead of proceeding with extended model versions.

Results from the Norwegian Logistics Model

The estimated coefficients and their corresponding selection criteria functions was implemented in the Norwegian Logistics Model by creating a new version of the executable ChainChoice program. This executable uses the same input data, but applies new selection criteria according to a logistic specification based on the estimated coefficients. The output of this executable is also slightly different; instead of one (deterministic) mode/weight combination for each freight flow, the model now outputs each potential (considered) choice alternative (i.e. each potential combination of mode choice and shipment size) as well as the predicted probabilities that these particular modes and shipment sizes are chosen jointly. Summing over the potential freight flows, multiplied by their corresponding probabilities for each mode, will therefore give the expected mode split predicted by the model. If, for each freight flow, the probabilities of all but one alternative are zero, the result of the stochastic model will be equivalent to the deterministic model.

We attempt to validate the model by comparing the deterministic model to the stochastic model through the demand elasticities for the three transport modes road, sea and rail, with respect to changes in time and distance-based link costs. The main findings are:

- All own elasticities and all cross elasticities for sea and rail have the expected signs. This is not the case for cross-elasticities for road transport, since road also is a part of the transport chains defined as “sea” and “rail”. For all commodities and both model types, changing the costs for sea will to a large extent lead to a mode shift to/from rail.
- For commodity groups 17 and 30 own elasticities in the stochastic case are consistently lower than the own elasticities in the deterministic case while there is no clear pattern for commodity group 13. We would expect the model to be less responsive.
- For virtually all commodity groups, all transport modes and both types of models, for both own elasticities and cross elasticities, the absolute value of the elasticity increases when the magnitude of the cost change increases. This relationship is stronger than expected.

Discussion and further work

Reducing the data set from the Swedish commodity flow survey to cover only the transport between Norway and Sweden, result in a minor number of commodities that fulfil the requirements regarding number of shipments and variation across transport modes and geographical areas. A possible way to validate the estimated choice models is to estimate the models using other commodity groups. A drawback with this approach is that most of the other commodity groups in the dataset have either too few observations and/or too little spread of transport modes and/or geographical areas to perform estimations. To increase the number of observations and the spread of mode choice and geography, the dataset from the Swedish commodity flow survey must either be increased to cover more than transport between Norway and Sweden, or be replaced with another data source. An alternative data source recently available is the Norwegian foreign trade statistics, that TOI has access to at shipment level, with detailed information about origin, destination, shipment size and value, and border crossing mode choice.

As discretising the choice of shipment size can be seen as a form of measurement error an alternative approach is to estimate a simultaneous discrete-continuous structural model with a joint choice of discrete transport mode and continuous shipment size.