"A Model for Freight Transport crossing Storebælt"

by
Anders Sørensen, COWI
Parallelvej 15, 2800 Lyngby
e-mail: ARS@COWI.DK

1 Introduction

In this note a model for growth in total freight transport crossing Storebælt is put forward. First, a model forecasts redistribution of freight generated by reduced transportation costs caused by the Fixed Link. Redistributed freight originates from present internal traffic within Jylland/Fyn or Sjælland/Lolland-Falster that will cross Storebælt in the future after establishment of the Fixed Link. This is the gravity model. Second, the results from the gravity model is compared with the existing literature.

In Section 2, the gravity model is analysed. Here, the applied data set is discussed and the model is formulated and estimated. Moreover, the model is applied to forecast redistributed traffic in relation to establishment of the Fixed Link crossing Storebælt. In Section 3, the estimated elasticities of freight volumes with respect to the production value are discussed and compared to estimates available from existing literature. Section 4 concludes.

2 Gravity model

In this section the gravity model that forecasts redistributed freight volumes generated by reduced transportation costs as a result of the Fixed Link is put forward. Redistributed freight volumes originate from the present internal traffic within Jylland/Fyn or Sjælland/Lolland-Falster that crosses Storebælt in the future after establishment of the Fixed Link. Hence, the calculated rise in freight volumes in the Storebælt passage when the Fixed Link has been established does not originate from new-generated traffic, but from traffic redistributed from other destinations.

The estimation and the estimation results are discussed in this section. The applied data set consists of observed freight volumes being transported between 15 zones. This implies that internal traffic within Jylland/Fyn or Sjælland/Lolland-Falster is known. In order to utilise this information redistributed traffic is estimated by a logit model, since this makes use of origin-destination combinations.

This section is composed of five sections. First, the data set used in the analysis is described. Second, the formulation of the model is discussed. Third, the applied model is chosen. Fourth, the estimation of the chosen model is discussed. Finally, forecasts of redistribution of traffic associated with lower transportation costs are performed.

2.1 Data

Data from different sources are applied in the analysis. The data set is divided in 15 geographical zones, corresponding to the Danish counties and the municipalities of Copenhagen and Frederiksberg. The county of Bornholm is excluded from the analysis due to lack of observations.

Data for domestic freight volumes with trucks between the 15 zones are available from the 'Kørebogsundersøgelse' obtained from Danmarks Statistik. In this data set, the aggregated freight volumes (in tonnes) and the number of trips are given for the first, the third, and the fourth quarter of

1995 and the first two quarters of 1996. The material is decomposed by type of cargo. The present analysis applies freight volumes classified by type of cargo.

Furthermore, logsums from a mode/route model for freight transport crossing Storebælt are applied. This variable is inversely related to transport costs between zones.

Production values (PV_j) and the gross domestic product in factor prices (BFI_j) from 1995 for the individual zones are available from Danmarks Statistik. Index j indicates zone j.

2.2 The model

In this section, the formulation of the applied gravity model is discussed. There is one model for each of three different types of goods: high-value general cargo, low-value general cargo, and bulk. In the following, the first model is referred to as model 1, the second as model 2, and the third as model 3.

Logsums are available for both low-value bulk and high-value bulk in the data set. This means that two models could be estimated for bulk instead of only one. However, the quality of logsums for low-value and high-value bulk are low, because the number of observations applied in the computation is relatively low.[1] Consequently, there appears to be small sample problems for low- and high-value bulk. These difficulties are less pronounced when one model for bulk is estimated, since the problems for both low-value and high-value bulk are dampened.

The formulation of the gravity model follows:

$$T_{ij}= \exp(\alpha \log(A_j) + \gamma \log(B_i) + \delta \text{ logsum}_{ij}),$$

where

- T_{ij} is the freight volume going from zone i to zone j.
- logsum_{ij} is a variable that is negatively related to the transport costs from zone i to zone j.
- A_j is a vector of attraction variables in zone j that measure the zonal demand for the type of freight modelled.
- B_i is a vector of generation variables in zone i that measure economic activity of the zone.

In the following, freight volumes from zone i are constrained, such that the total freight volume out of zone i ($\Sigma_{j'} T_{ij'}$) equals the observed value. This implies that the model is single-constrained. A consequence of the constraint is that the generation variable, B_i, becomes irrelevant in the model formulation.

$T_{ij}/\Sigma_{j'} T_{ij'}$ is the probability for a truckload of goods to go from zone i to zone j. By using the formulation of the gravity model it can be shown that these probabilities can be formulated as a multinomial logit model. This formulation is used for estimation. An observation in terms of the logit

[1] 5,866 and 2,934 observations enter the computation of logsums for high-value and low-value general cargo, respectively. The corresponding numbers are 1,631 for high-value bulk and 1,043 for low-value bulk.

model thus consists of a number of tonnes being transported from zone i to zone j. The probability of observing this choice is $T_{ij}/\Sigma_{j'}T_{ij'}$ and T_{ij} is the weight of the observation.

The logit model is based on the idea that behaviour of a person or a company is to choose the utility maximising alternative among a set of choices. That is, the probability for a person to choose a specific alternative equals the probability for this alternative to maximise utility. Therefore, in this framework a utility function is specified:

$$U_j = V_j + \varepsilon_j$$

$$= \beta_1 \text{ logsum}_{ij} + \beta_2 \text{ inter}_{ij} + \beta_3 \log(PV_j) + \beta_4 \log(BFI_j) + \varepsilon_j.$$

Total utility, U_j, is composed of a deterministic term, V_j, and a random component, ε_j. The disturbance terms are assumed to be independent and identically Gumbel distributed.

The explanatory variables are:

- logsum_{ij}, see above;
- inter_{ij}: an east-west dummy. $\text{inter}_{ij}=1$ for traffic crossing Storebælt, i.e. $i\in(8,15)$ and $j\in(1,6)$ or $i\in(1,6)$ and $j\in(8,15)$. Otherwise $\text{inter}_{ij}=0$;
- PV_j: Production values in zone j;
- BFI_j: Gross domestic production in zone j.

Logsums for traffic crossing Storebælt are calculated differently than non-crossing traffic. Therefore, a dummy, inter_{ij}, is introduced to capture the effect arising from the different calculation methods.

Production values and gross domestic productions are attraction variables of zone j according to the formulation of the gravity model. In the logit set-up the variables measure the size of the different zones. Therefore, these variables enter the model in logarithms, i.e. $\log(PV_j)$ and $\log(BFI_j)$ are included in the model formulation.

The utility effect from logsum_{ij} is expected to be positive, since a higher logsum implies lower transportation costs between zone i and zone j. Furthermore, a higher level of economic activity in a zone - measured by higher levels of $\log(PV_j)$ and $\log(BFI_j)$ - is expected to imply a higher level of utility. Therefore, β_1, β_3, and β_4 are expected to be positive. We have no expectation to the sign of β_2 a priori.

A large share of the traffic in the counties is internal. It seems unreasonable that more than a small proportion of this traffic could be redistributed to go east-west after establishment of the Fixed Link. Furthermore, if these observations were used in the analysis the model behaviour could become unbalanced such that improbable volumes of traffic could be assigned to the Fixed Link. Therefore, observations with same origin and destination and the corresponding alternatives are excluded from the analysis.

Two versions of the logit model have been estimated. In the first version destination is applied as alternative in the model formulation. This implies that cargo can be transported to 13 different zones. In the second version of the model origin is alternative, while observations are identified by destination. Again, one zone out of 13 must be chosen. The first version turned out to give best results and is applied in the following.

2.3 Estimation

In this section different variations of the model are estimated. The most appropriate version, is found and applied in the sections below.

A multicollinearity problem exists, since $\log(PV_j)$ and $\log(BFI_j)$ are highly correlated ($\rho_{\log(BFI),\log(PV)} = 0.996$). This shows up in the regressions as negative and/or insignificant estimators of β_4. In an attempt to solve the problem a new variable, RES_j, has been constructed. RES_j is the estimated error term from:

$$\log(BFI_j) = \gamma_0 + \gamma_1 \log(PV_j) + RES_j.$$

This regression is applied because the part of $\log(BFI_j)$ that cannot be explained by $\log(PV_j)$ is given by RES_j. Consequently, the information included in $\log(BFI_j)$ and $\log(PV_j)$ is the same as the information included in RES_j and $\log(PV_j)$. The advantage, however, of applying the latter pair of variables is that RES_j and $\log(PV_j)$ are uncorrelated by construction, implying that the model includes no multicollinearity problem.

The reason why the relation between BFI_j and PV_j is formulated in logarithms is that the variables are size measures for zones. Hence, the lower the values of BFI_j and PV_j, the smaller is the size of zone j. This implies that the part of BFI_j that is not explained by PV_j is positively related to PV_j. In other words, the variance of the error term in the regression increases with PV_j. This is taken into account by applying the above formulation.

According to the above discussion

$$V_j=\beta_1 \text{logsum}_{ij} + \beta_2 \text{inter}_{ij} + \beta_3 \log(PV_j) + \beta_4 RES_j$$

is estimated. The expected sign of β_4 is positive because a higher gross domestic product results in an increase in RES_j and, thereby, in a higher level of utility. However, the point estimates of β_4 turn out to be negative and/or insignificant.

A similar estimation is performed for $\log(BFI_j)$ as dependent variable and $\log(PV_j)$ as explanatory variable. Again, the estimated coefficients to the error term do not improve the explanatory power of the models implying that the statistical quality of the model is not improved by including the variable RES_j. The explanation for RES_j entering the model insignificantly is the very high correlation between $\log(PV_j)$ and $\log(BFI_j)$. Hence, the extra information obtained in the model by including RES_j is small.

As a consequence of the above discussion only $\log(PV_j)$ is included in the model. Therefore,

$$V_j=\beta_1 \text{logsum}_{ij} + \beta_2 \text{inter}_{ij} + \beta_3 \log(PV_j).$$

That is, $\log(PV_j)$ enters the model, while $\log(BFI_j)$ is excluded. The reason is that the demand for transportation services increases through two channels when the production value increases. First, the volume of trade increases implying that more final goods have to be transported to the market place. Second, more raw materials have to be transported to the production side of the economy. Since the production value is composed of raw materials besides value added, the latter effect is included in the model when $\log(PV_j)$ is applied. Hence, economic activity is believed to be more relevant to freight transportation than value added. Furthermore, the statistical quality of the model including $\log(PV_j)$ is better than the model including $\log(BFI_j)$ measured by the goodness-of-fit.

2.4 Results

In this section, the chosen model is estimated and discussed. In Table Fejl! **Ingen tekst med den anførte typografi i dokumentet..1** the estimated coefficients and t-statistics (in brackets) for β_i, i=1,2,3, in the three models are given. Furthermore, two goodness-of-fit measures are shown.

Table ***Fejl! Ingen tekst med den anførte typografi i dokumentet.****.1 Coefficients for the estimated models (t-statistics in brackets)*

	β_1	β_2	β_3	ρ^2 w.r.t. zero	ρ^2 w.r.t constants
Model 1: high-value general cargo	0.2637 (15.7)	-0.6327 (-6.7)	0.7774 (25.7)	0.1574	0.1266
Model 2: low-value general cargo	0.6182 (23.0)	-0.4838 (-4.3)	0.3302 (9.7)	0.1663	0.1429
Model 3: bulk	0.9862 (28.9)	-1.188 (-11.9)	0.1100 (3.7)	0.2710	0.2606

The fit is generally good and the models are significantly better than models only including constants. It appears that the effect on utility from $logsum_{ij}$ is positive. This is also the case for $\log(PV_j)$. The dummy affects utility negatively. All coefficients are highly significant.

To evaluate the significance of the estimated coefficients two groups of changes are performed. First, PV_j is increased/lowered by 10% for the three/seven/eleven zones with the highest PV_j. Second, the same changes are performed for $logsum_{ij}$. In the first group of changes the estimated values of β_1 are within an interval of ±2% around the baseline value of β_1 from Table Fejl! **Ingen tekst med den anførte typografi i dokumentet..1**. Similarly, the intervals for β_2 and β_3 are ±5% and ±10% around the baseline values, respectively. In the second group of changes, the intervals for β_1 and β_2 are within ±12.5% and ±10% around the baseline values. The interval for β_3 equals ±15% for model 1 and model 2, but is as high as ±75% for model 3. This implies that the elasticity of freight transportation of bulk with respect to the production value is sensitive to changes in logsums for bulk.

In general, point estimates in different models cannot be compared directly. The reason is that the estimated coefficients actually equal β_1/μ, β_2/μ, and β_3/μ, where μ is related to the variance of the error term. Consequently, the estimated coefficients have to be normalised or μ must have the same value in the models under comparison.

According to the goodness-of-fit it seems reasonable to assume that μ attains roughly the same value in model 1 and model 2. By comparing the coefficients for the two models two characteristics appear. First, freight volumes of low-value general cargo are more sensitive with respect to logsums than high-value general cargo. This result is in accordance with intuition: Transportation costs constitute a higher share of the value of low-value goods than high-value goods. Therefore, the freight volumes of low-value goods are more sensitive to changes in logsums than high-value goods.

Second, freight volumes of high-value general cargo is more sensitive to changes in $\log(PV_j)$ than freight volumes of low-value general cargo. This characteristic can be explained by an effect from the household side of the economy. The intuition is that goods such as furniture, clothing etc. are more sensitive to changes in income than goods such as food. In the data set, the classification of goods with high income elasticity tend to be high-value goods, whereas goods with low income elasticity tend to be low-value goods. Hence, the consumption of both high-value and low-value goods increase when the production value and, thereby, income increases. However, the share of high-value goods in the composition of consumption increases, which makes freight of high-value goods more sensitive to changing production value than freight of low-value goods.

It seems reasonable that the effect on freight volumes of bulk from $\log(PV_j)$ is below or equal to the effect from $\log(PV_j)$ on freight volumes of low-value general cargo, i.e. β_3(bulk) $\leq$ β_3(low-value general cargo). The reasoning is that the characteristics of some of the commodities in the two groups seem to be close in nature. Furthermore, on average, the goods classified under bulk seem to be less sensitive to changes in the production value compared to the goods classified under low-value general cargo.

β_3 attains the same value in model 2 and model 3 if freight volumes of low-value general cargo and bulk are equally sensitive to changes in $\log(PV_j)$. This implies that β_3 is used to normalise the estimated coefficient for logsums, β_1, that is, $(\beta_1/\mu)(\mu/\beta_3)= \beta_1/\beta_3$ is used to evaluate the sensitivity of freight volumes for bulk and low-value general cargo. It is evident from Table Fejl! **Ingen tekst med den anførte typografi i dokumentet..1** that β_1/β_3 equals 1.87 for low-value general cargo and 8.97 for bulk. Hence, freight volumes of bulk is far more sensitive to changes in logsums than low-value general cargo. According to the above discussion high-value general cargo is less sensitive to changes in logsums than high-value general cargo and, therefore, also bulk.

Bulk is even more sensitive to changes in logsums than both types of general cargo if freight volumes of bulk is less sensitive to changes in $\log(PV_j)$ than freight volumes of low-value general cargo, that is, if β_3(bulk) < β_3(low-value general cargo). The intuition for these results follow the idea that transportation costs constitute a higher share of the value of bulk than is the case for both high-value and low-value general cargo.

2.5 Redistributed Traffic

In this section, forecasts of redistributed traffic associated with lower transportation costs are performed.

In order to forecast redistributed traffic in relation to establishment of the Fixed Link an experiment is carried out. When the Fixed Link is opened, relative transportation costs are expected to change such that traffic crossing Storebælt becomes cheaper compared to traffic not crossing Storebælt. Hence, in the experiment transportation costs for crossing the Storebælt passage are lowered, while the costs for non-crossing traffic remain unchanged.

In the following, $logsum_{ij}$ for traffic crossing Storebælt is increased, whereas $logsum_{ij}$ for internal traffic in Jylland/Fyn or Sjælland/Lolland-Falster is unchanged. The change in logsums is calculated by lowering the costs all ferry services by 100 DKK.

Table Fejl! **Ingen tekst med den anførte typografi i dokumentet..2** shows the percentage change in the traffic crossing Storebælt generated by the fall in ferry rates. For example, the volume of high-value general cargo crossing Storebælt increases by 2.8% when ferry rates are lowered by 100 DKK.

Table ***Fejl! Ingen tekst med den anførte typografi i dokumentet.****.2 Change in Storebælt crossing traffic from falling ferry rates*

	Percentage change
Model 1: high-value general cargo	2.80
Model 2: low-value general cargo	4.73
Model 3: bulk	4.91

It appears from Table Fejl! **Ingen tekst med den anførte typografi i dokumentet..2** that the changes in traffic crossing Storebælt follow the sensitivity results discussed in Section 2.4. This is the case even though the percentage change in transport costs does not attain the same value for all three types of cargo.

3 Evaluation of Elasticities

In this section the effect on freight volumes from economic growth - measured by the elasticity of freight volumes with respect to production value - is discussed.

3.1 Existing literature

This section discusses estimates of the elasticity of freight volumes with respect to production value for Denmark from the existing literature. The only source, however, known to us is Bjørner (1997).

Bjørner (1997) estimates the elasticity of transport (tonnes kilometres) and the elasticity of traffic (kilometres) both with respect to the production value in fixed prices in a time series analysis on macro-data by using the "Johansen Procedure". The point estimates of the two long-run elasticities are 1.32 and 0.92, respectively.

The ratio of transportation (tonnes kilometres) to traffic (kilometres) is a measure of the average utilisation of trucks.[2] Consequently, the difference of the elasticity of transport and the elasticity of traffic with respect to the production value equals the elasticity of average utilisation of trucks with respect to the production value. This elasticity, however, equals the elasticity of freight volume with respect to the production value if the estimated elasticities also apply at the micro level. The elasticity of the average utilisation of trucks with respect to the production value is evaluated to be a rough estimate of the elasticity of freight volume with respect to the production value.

According to the above discussion, the two estimated elasticities are applied to calculate an rough estimate of the elasticity of freight volume with respect to production value. This implies

elasticity of freight volume w.r.t. production value

≈ elasticity of transport w.r.t. production value

- elasticity of traffic w.r.t. production value.

Hence, according to this investigation the elasticity of freight volume with respect to the production value is around 0.4.

3.2 The present analysis

In this section, the estimated elasticities from section 2.4 are compared to the value given in the previous section.

The estimates of β_3 from Section 2.4 can be interpreted as the elasticity of freight volume with respect to the production value. The justification follows from the fact that the model applied in this note is a gravity model, see Section 2.2. The average elasticity equals 0.41.

Consequently, the elasticities of freight volumes with respect to the production value estimated in Section 2.4 are evaluated to be reasonable. This is because the average value obtained in the present cross-section analysis is close to the value achieved in the time-series analysis by Bjørner (1997) given that the estimated elasticities are valid at the micro level.

4 Conclusion

In Section 2 of this note a gravity model for freight transport crossing Storebælt is formulated and estimated. Two explanatory variables enter the models. This is logsums that is a variable negatively related to transport costs, and the logarithms to regional production values that is a measure of economic activity in the individual zone. Three models - identified by low-value general cargo, high-value general cargo, and bulk - are estimated. The estimated logit models fit the data reasonably. Furthermore, all of the signs for the estimated coefficients are in accordance with expectations. In addition, all variables enter the models significantly.

In Section 3 different estimates of the elasticity of freight volumes with respect to the production value are discussed. It is found that the estimated elasticities in the present investigation correspond well with comparable elasticities for Denmark in the existing literature.

---

[2] The ratio is a weighted average of truck loads, where the weights are the length of the single trip out of total traffic.

References

"Demand for Freight Transportation and Freight Traffic" by Thomas Bue Bjørner, SØM Publikation nr. 17, AKF Forlaget.