A Commuting Model[1]

Anne Kaag Andersen
AKF, Institute of Local Government Studies - Denmark
Nyropsgade 37
DK-1602 Copenhagen V
Denmark

Phone: +45 3314 5949 ♪ 74
Fax: +45 3315 2875
E-mail: aka@akf.dk

1. Introduction

In this paper two different versions of a commuting model applicable for a regional macroeconomic model are presented. The first is a gravity model type, while the second is a logit model. The model types are presented, and estimation results are given. First however, data are presented.

2. A Commuting Model for LINE

LINE is a macroeconomic interregional model at municipal level for Denmark. It is being developed at AKF, Institute of Local Government Studies – Denmark. It can be classified as a model concentrating on the economic activity and the economic flows between the Danish municipalities. Traffic is not modelled very detailed, but the money flows for commuting, shopping, trade and tourism form an excellent basis for analyses of traffic flows of different kinds. All activities in the economy are classified according to the locality as well as to the standard subdivision in *Social Accounting Matrices* (SAM), i.e. into groups of sectors, qualification groups, institution components and commodities. The different flows in LINE are illustrated in Figure 1.

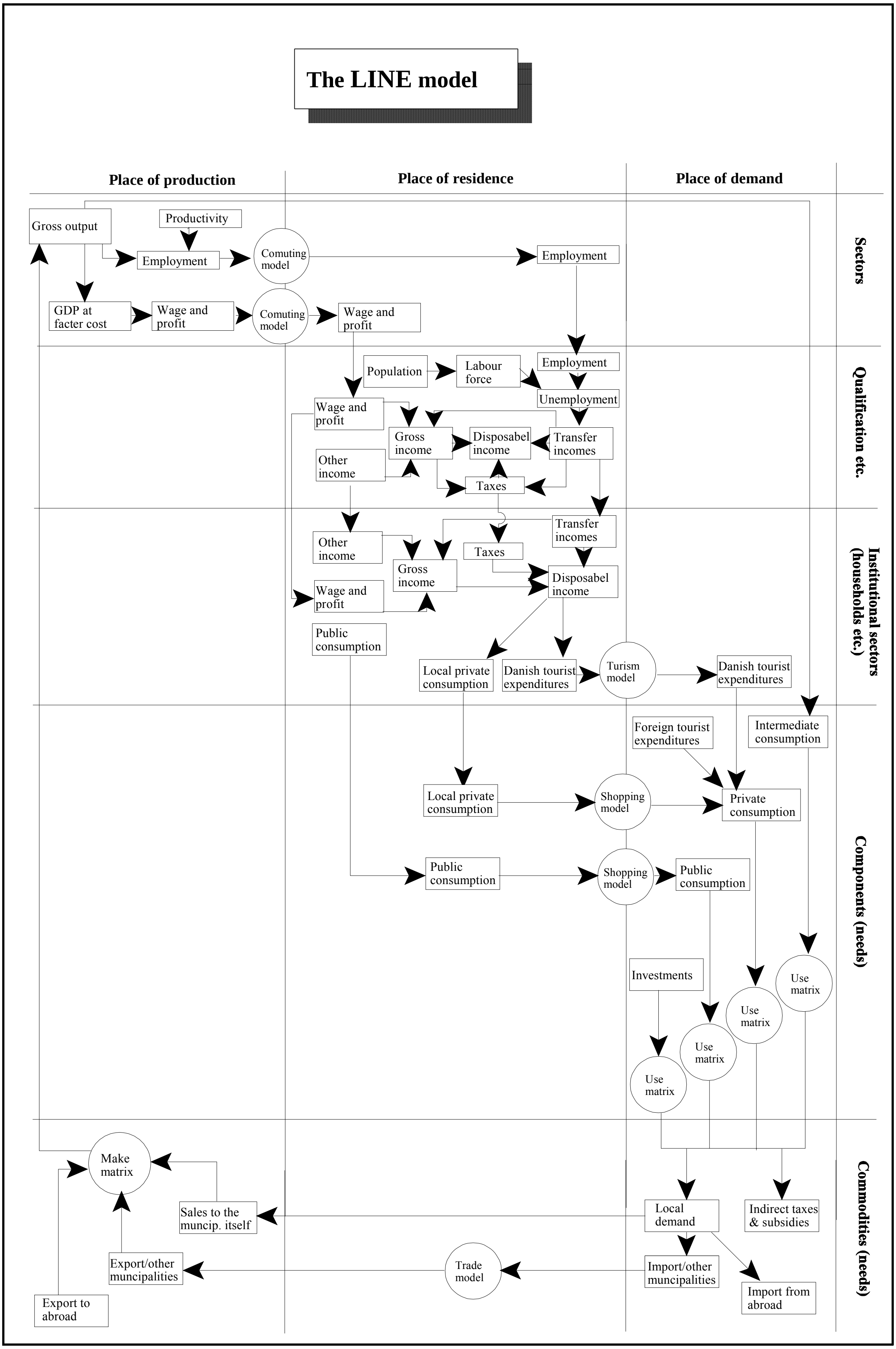
It is the intention to expand the model to include a cost circle, running opposite the real circle as illustrated in the figure. Furthermore, it is the plan to change the flow in the upper right corner, in order to let commuting be determined by groups instead of sectors. In that way it is possible to model the labour market endogenously, also including migration as well as births and deaths. For a detailed description of LINE see Filges and Rasmussen (1998) or Madsen et al. (1997).

In this paper, two different versions of a possible submodel for commuting in LINE are presented. First, focus is upon data, however.

---

[1]This paper is a compilation of parts of two papers from my PhD-thesis handed in at Institute of Economics, University of Copenhagen, April 1999 (Andersen 1999a and Andersen 1999b).

Figure 1: The LINE model



3. Data

The data used in the analyses of commuting stem from Statistics Denmark. From registers the addresses of residence and workplace are known for all employed. The data are aggregated to municipality level for each sector. It is not the number of commuting *trips* between two municipalities which is modelled, but the number of employees living in a specific municipality and working in another. Some individuals live far away from their workplace. It is not probable that these individuals commute twice a day, they probably only commute in weekends. Furthermore, some employees are registered in another municipality than they actually work in, as an example some artisans can be mentioned. For a further discussion, see e.g. Miljøministeriet (1994). The data used in this analysis are for 1995.

To obtain a measure of spatial separation between the municipalities in Denmark a computer programme is built up in *Pascal*. It divides Denmark, as well as a minor part of Sweden, into eight parts. Within each of these parts a journey can be made crossing land only. Data on distances following the road network stem from Vejdirektoratet (1994). For distances between municipalities which are separated by water a more complex routine has to be used. It is assumed that individuals choose the cheapest/shortest route. To find this, all possible routes are evaluated.

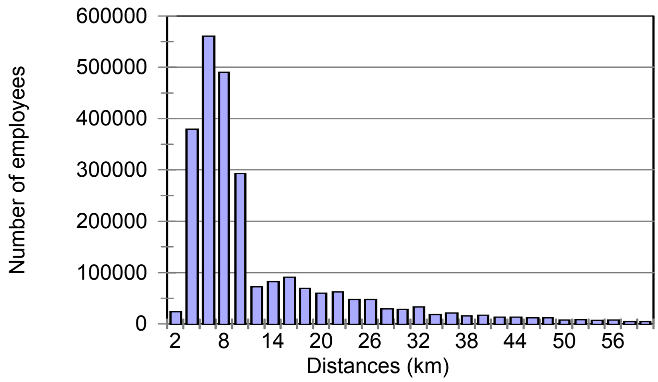
The resulting matrix is at municipality level. There are some general problems concerning neighbouring municipalities. Between those it is possible to have a much shorter distance, than the one measured. Besides, there are problems concerning distances *inside* the municipalities. They can be set to zero, or another standard value, or they can be assumed dependent on the size of the municipality. Frost and Spence (1995) discuss the problem. Here, intra municipality distances are set to 2/3 times the square root of the radius for a circle with the same area as the municipality. This is the mean distance of trips to the centre for origins evenly spread out at the circle.

When calibrating the behavioural models the measure of spatial separation should reflect the inconvenience experienced by a commuter. For that purpose, a matrix of commuting *costs* is used. Distances by land are weighted by the costs per kilometre driving a car, while for distances by sea, the price for bringing a car with a driver is used. The cost per kilometre is, with a rough estimate, set to 1 DKK (Rasmussen 1996). Furthermore, the costs are modified according to the Danish tax rules for deductions for commuters. The resulting costs matrix fits car traffic only. This is of course not fully adequate to all means of transportation, but can be used as a proxy for the general travel costs.

When discussing necessary and wasteful commuting in terms of environmental problems, *distances* are more relevant. Distances by land and by sea are treated equally. It could be argued, that distances by sea should be weighted higher due to higher consumption of energy per person when using a ferry than when driving a car (cf. Trafikministeriet 1996). This is ignored here, however.

The average commuting distance using the distance matrix as defined above, is given by 15.8 km. As illustrated in Figure 2, the typical commuting distance is somewhat smaller. Only individuals with commuting distances below 60 km are included in the figure. There is, however, a considerable number of individuals with rather long commuting distances, enhancing the average commuting distance. It should be noticed also that especially the short commuting distances are encumbered with uncertainty, due to the ad hoc definition of intra-municipality distances.

Figure 2. The distribution of commuting distances



Note: The observations are grouped into intervals of two kilometres. Only observations less than 60 km are included.

In some municipalities the number of workplaces is higher than the number of employees living there, while the opposite is true in other municipalities. In urban municipalities, the number of workplaces is typically higher than the number of employees living there. The number of employees working in a municipality, can be denoted the *day population,* while the number of employees living in a municipality can be denoted the *night population.* The absolute differences between day population and night population for the Danish municipalities are illustrated in Figure 3. It is seen that there are huge differences between day and night population. The area of Copenhagen has a high number of workplaces, and attracts employees from other areas. The same is true for several cities, e.g. Århus, Odense, Randers.

For this given location of households and workplaces there is a minimum required commuting, which is necessary in order to secure the number of employees respectively living and working in the municipality, i.e. the night and day population. The necessary commuting is found by minimizing total travel costs, given constraints securing the marginal sums. The necessary commuting can be compared with the *actual commuting*, giving *wasteful commuting* as a residual. The term wasteful refers to a normative approach, where it is the objective to minimize the commuting distance, since it is waste. Even though this statement can be questioned, it is certainly relevant to constrain the actual commuting, given the environmental problems caused by transport. The amount of wasteful commuting gives a hint on the causes for the actual commuting.

In Table 1 the average and total amounts of these concepts are shown. The above described distance matrix is used. The wasteful commuting constitutes 56% of the total amount of commuting.

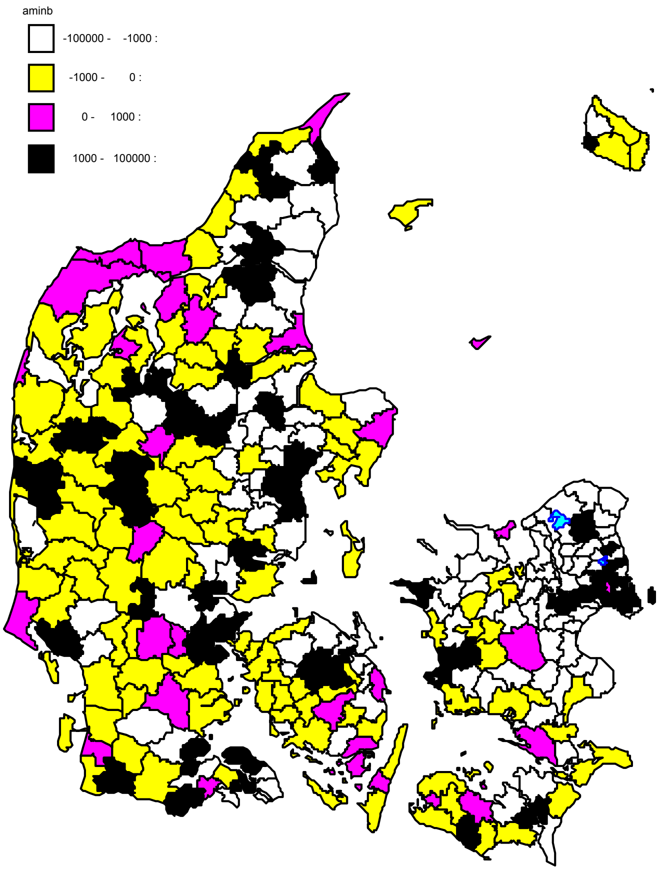


Figure 3. Day population minus night population

The wasteful commuting exists for many reasons: only few jobs and houses are available at a specific point in time; people are married to each other and by that not able to choose any specific residence location; and the housing market as well as the labour market are heterogenous. The latter implies that jobs as well as houses are not fully substitutable. If it is the objective to reduce the amount of commuting, different policies should be taken into account in order to influence the two different types of commuting. To reduce the necessary commuting, the location of houses and workplaces should be influenced, but to reduce the wasteful commuting, individuals' behaviour when choosing job and residence should be influenced. The latter could for example be done by increasing commuting costs, while the former could be done via physical planning.

Table 1. Actual, necessary and wasteful commuting respectively for Danish municipalities 1995

	Actual	Necessary	Wasteful
Total (kilometres)	41,607,883	18,387,220	23,220,663
Average (kilometres per employee)	15.8	7	8.8

For estimation of the logit model, several other variables for the municipalities are used too. These stem from different sources, e.g. *Indenrigsministeriets Nøgletal*.

4. Commuting models

It is the employment per workplace, residence and sector, q_{abe} which is under consideration, or more precisely: it is the objective to analyse the quotients transferring q_{ae} (employment per workplace and sector) to q_{be} (employment per place of residence and sector). In this paper, the sector dimension e will be suppressed; but all the models can be applied for the respective sectors.

Let O_a be the marginal sum of q_{ab} over b, i.e. the day population, and let D_b be the marginal sum of q_{ab} over a, i.e. the night population. Furthermore, let c_{ab} be the cost matrix.

Below, two different model types are presented: the gravity model and the logit model. Both of them applies directly to the commuting situation.

4.1. Spatial Interaction Models

In many models for residential location and commuting, spatial interaction models play an important role. A standard representative of these is the gravity model. Next follows a presentation of the gravity model.

There are different types of the gravity model. The model can be doubly constrained, singly constrained or unconstrained. By using the doubly constrained model, the number of workplaces as well as individuals living in each municipality are assumed to be given. By using the singly constrained it is either assumed that the number of workplaces or the number of individuals living in each municipality is given. Using the unconstrained neither of them is given. Even though one or both of the factors are unknown, there are factors, or proxy variables, describing the attraction of each municipality. Here, a doubly constrained gravity model is looked upon.

The doubly constrained gravity model can be formulated as:

$$q_{ab} = A_a B_b O_a D_b \exp(-\beta c_{ab}) \qquad \textbf{(1)}$$

where A_a and B_b are so-called balancing constants, and are given by:

$$A_a = [\sum_b B_b D_b \ \exp(-\beta c_{ab})]^{-1} \qquad \textbf{(2)}$$

$$B_b = [\sum_a A_a O_a \ \exp(-\beta c_{ab})]^{-1} \qquad \textbf{(3)}$$

As can be seen, every A_a is a part of the expression for B_b and vice versa. It is not possible to find an analytical solution, and it is therefore necessary to determine the balancing constants iteratively.

Intuitively, the model seems reasonable. The interaction between *a* and *b* is positively correlated with the number of workplaces in municipality *b* and the number of residents in municipality *a.* Similarly, the interaction is negatively correlated with the level of travel costs between the two municipalities.

A slightly different gravity model exists too:

$$q_{ab} = A_a B_b O_a D_b c_{ab}^{-\gamma} \tag{4}$$

The distance function in the balancing constants is changed accordingly. The choice of model should in principle be based on knowledge of the cost constraint. In practice, however, the characteristics of the resulting distance function are used as criteria.

The distance function describes how sensitive the number of trips is to the cost of travel. A difference between the two formulations is that while the parameter of the power function is dimension less, the parameter in the exponential function depends on the dimension of the costs included, as illustrated in Fotheringham and O'Kelly (1989). If the distance parameter (β) approaches infinity, the result of the gravity model equals the minimum or necessary commuting, as touched upon in section 3.

4.2 The Logit Model

The logit model can be derived via stochastic utility functions. For a thorough derivation of the logit model, see for example Anas (1982) or Ben-Akiva and Lerman (1985). In the situation looked upon here, a utility function is assumed for the segment of individuals who work in municipality *a*. The utility obtained by living in municipality *b* is given by

$$U_a(b) = V_a(b) + \mu_{ab} \tag{5}$$

where $V_a(b)$ is the part of the utility function, which is deterministic with exogenous variables as arguments, while μ_{ab} is a stochastic part. By assuming that the stochastic parts are independent and identically distributed with the Weibull distribution, and that individuals choose the alternative which yields the highest utility, the well-known logit model comes up:

$$p_a(\bar{b}) = \frac{\exp(V_a(\bar{b}))}{\sum_b \exp(V_a(b))} \tag{6}$$

where $p_a(\bar{b})$ is the probability of choosing the specific municipality $\bar{b}$ for residence for an individual, who works in municipality *a*. The deterministic part of the utility model, $V_a(b)$ is assumed to be a linear function of several characteristics of the choice, for example commuting costs and local tax level. The parameters in the utility function can be assumed to be equal for all workplace municipalities, i.e.:

$$V_a(b) = \sum_{i=1}^{I_1} \beta_i x_{ib|a} + \sum_{i=I_1+1}^{I} \beta_i x_{ib} \tag{7}$$

Here, the first I_1 variables depend on the municipality chosen for residence as well as the workplace municipality, an example is the commuting costs. The remaining variables depend on the residential municipality only, an example is the amount of forest.

No constant is included in the utility function, since terms that do not differ between the alternatives fall out of the probability (cf. Greene 1990). Constants varying with the alternatives can be included. If enough alternative specific constants are included, the model becomes doubly constrained in the sense that the estimated probabilities will equal the observed. An example using this approach is given in Berglund and Lundqvist (1998).

4.3 The Logit model and the gravity model

The gravity model is formulated in *numbers* of individuals interacting between two zones. The logit model is formulated as *probabilities* of making a certain decision for individuals in a specific group. There is, however, a close connection between the two types of models. As shown by Anas (1983), proper definitions of utility functions can transform a logit model to the unconstrained, singly constrained or the doubly constrained gravity model respectively.

Not only the models are alike; the estimates obtained from the logit model via maximum likelihood estimations correspond to the values of A_i, B_j and β, obtained when the standard iterative procedure is used for the gravity model as demonstrated by Anas (1983). This can be illustrated by looking at the first order conditions for the maximum likelihood problem for the logit model. A similar proof for the gravity model, assuming a multinomial distribution, is given by Evans (1971).

5. Empirical Results

For all employees aggregated the resulting doubly constrained gravity models are given by:

$$q_{ab} = A_a B_b O_a D_b \exp(-0.12 c_{ab}) \qquad \textbf{(8)}$$

with associated balancing constants, and

$$q_{ab} = A_a B_b O_a D_b c_{ab}^{-2.41} \qquad \textbf{(9)}$$

with associated balancing constants, respectively. The c_{ab} matrix is the cost matrix used for calibration.

Different variables explaining individuals' choice of residential location is chosen for estimation of the logit model. The result is shown in table 2, with corresponding *t*-values in parentheses.

The signs of the parameters are as expected; negative for commuting costs, housing costs, tax level and share of social security recipients, and positive for the dummy variable and share of forest. It is difficult to judge about the size of the variables directly. Instead it is fruitful to look at elasticities and trade-offs. The elasticities are shown in table 2 as well. The parameter sizes are influenced by the choice of variables included, and should therefore be interpreted with care.

Table 2. Estimation results and elasticities, logit model

Variable	Unit	Logit model	Elasticities
Commuting costs	100 DKK	-5.04 (-917)	-0.62
Dummy	-	2.74 (1415)	---
Housing costs	100 DKK	-0.53 (-417)	-1.19
Forest	%	0.01 (118)	0.07
Tax	%	-0.08 (-90)	-1.39
Social security recip.	%	-0.13 (-288)	-0.79
Pseudo R^2		0.59	
ρ^2		0.69	

It is important to include enough explaining variables. Lack of data is a major reason for exclusion of relevant variables. Anas (1982) lists ten major sources of bias in an estimated model. Among these are the specification of the choice model, the specification of the utility function (form and variables) and data measurement methods. The selection of the explaining variables is characterised as an important source of bias. Exclusion or misspecification influences the error terms. To judge whether the model is valid or not, it is possible to look upon elasticities and trade-offs, as well as to calculate some measures.

As is seen from the t-values all variables are judged to be extremely significant. The very high t-values are due to the high number of observations. To validate the model, two measures of fit are reported, i.e. *pseudo* R^2 and ρ^2.

6. Conclusion

Two different versions of a commuting model applicable for the LINE model have been presented. The models can be used in LINE, for policy experiments as well as for extrapolations. Further discussion of the models, extensions etc. are of course possible. Some of these are given in Andersen (1999a) and Andersen (1999b).

It is furthermore the plan to combine models for commuting with models for migration, some first thoughts are given in Andersen and Filges (1998).

7. References

Anas, A. (1982): *Residential Location Markets and Urban Transportation. Economic Theory, Econometrics, and Policy Analysis with Discrete Choice Models*. Academic Press.

Anas, A. (1983): Discrete choice theory, information theory and the multinomial logit and gravity models, *Transportation Research,* vol 17B, pp 13-23.

Andersen, A.K. (1999a): Modelling Commuting and Residential Location – Spatial Interaction Models, in *Location and Commuting*, PhD Thesis.

Andersen, A.K. (1999b): A Commuting Model for the Danish Municipalities. In: Andersen, A.K.: *Location and Commuting*, PhD Thesis.

Andersen, A.K. and T. Filges (1998): *Migration and commuting patterns of the Danish Population.* Paper presented at TLEnet workshop, Pärnu, Estland, October 1998.

Ben-Akiva, M. and S. Lerman (1985): *Discrete Choice Analysis.* MIT Press.

Berglund, S. and L. Lundqvist. (1998): Barrier in Travel models, in S. Berglund: *Analysing Transportation Patterns in Cross-Border Contexts.* Licentiate thesis, Kungl. Tekniska Högskolan, Stockholm.

Evans, A.W. (1971): The Calibration of Trip Distribution Models with Exponential or Similar Cost Functions. *Transportation Research*, vol. 5, pp. 15-38.

Filges, T. and J.L. Rasmussen (1998): LINE. A Model of 275 Municipalities, Paper presented at *Structures and Prospects of Nordic Regional Economies,* a Research Seminar of the Nordic Section of the European Regional Science Association, June 1998.

Fotheringham, A.S. and M.E. O'Kelly (1989): *Spatial Interaction Models: Formulations and Applications.* Kluwer Academic Publishers.

Frost, M.E. and N.A. Spence (1995): The rediscovery of accessibility and economic potential: The critical issue of self-potential. *Environment and Planning A*, vol. 27, pp.1833-1848.

Greene, W.H. (1990): *Econometric Analysis.* MacMillan.

Madsen, B.; C. Jensen-Butler and P.U. Dam (1997): *The LINE-model*. AKF, draft unpublished.

Miljøministeriet (1994): *Pendlingen i Danmark.*

Rasmussen, S. (1996): Dyrere at holde bil. *Motor*, no. 1, pp 28-33.

Trafikministeriet (1996): *TEMA – En Model for transporters emissioner. Dokumentationsrapport*.

Vejdirektoratet (1994): *Vejafstande i Danmark*.