Spatial modelling of origin-destination commuting flows in Switzerland

Thomas Schatzmann*
Institute for Transport Planning and Systems, ETH Zurich
HIL F 32.1, Stefano-Franscini-Platz 5, 8093, Zurich, Switzerland
E-Mail: thomas.schatzmann@ivt.baug.ethz.ch

Georgios Sarlas
Institute for Transport Planning and Systems, ETH Zurich
HIL F 51.3, Stefano-Franscini-Platz 5, 8093, Zurich, Switzerland
E-Mail: georgios.sarlas@ivt.baug.ethz.ch

Kay W. Axhausen
Institute for Transport Planning and Systems, ETH Zurich
HIL F 31.3, Stefano-Franscini-Platz 5, 8093, Zurich, Switzerland
E-Mail: axhausen@ivt.baug.ethz.ch

* Corresponding author

Submission Date for the abstract: March 15 2018

Submitted for presentation at hEART 2018 – 7th Symposium of the European Association for Research in Transportation, September 5-7, 2018, Athens, Greece

Abstract

Objective and methodology

This paper presents a direct modelling approach for origin-destination public transportation commuting flows with possible endogenous regressors for the case of Switzerland. Its purpose is to improve the gravity modelling approach for OD flows by applying a spatial autoregressive regression model and testing different spatial weighting schemes. To the best of our knowledge, there has been no prior application of such advanced models in the context of transport demand modelling for public transport. Methodologically, in the first step a gravity model is developed and tested for the presence of spatial autocorrelation in its residuals. Subsequently, variants of a spatial lag model with different spatial weighting schemes are developed. Furthermore, we test a variable based on mean income differences on its ability to describe interregional demand patterns. In addition, we treat for the endogenous nature of the newly constructed variable. We are also testing its ability to serve as the basis for the construction of the spatial weight matrix, thus replacing the commonly used travel time / distance metric. On the modelling front, we use an Ordinary least squares (OLS) estimator for the gravity model, while a Generalized Method of Moments and Instrumental Variable (GMM/IV) (IV) estimator for the spatial models is employed in order to obtain unbiased and consistent parameter estimates. Last, we evaluate various models' goodness-of-fit measures and in-sample predictive accuracies by comparing among each other as well as to those of a state-of-the practice transport model (as provided by national spatial planning bureau (NPVM)). This comparison can allow us to draw solid conclusion with respect to the suitability of the presented method for predicting commuting flows.

Case study

In brief, a case study for public transport commuting flows in Switzerland is designed to illustrate the concept of OD flow modelling, based on travel-to-work trips data from the Federal Census of 2000. The data cover 2896 Swiss municipalities and contains over 250'000 observations on their initial form. However, it does not fill the whole flow matrix that contains 2896^2 = 8,386,816 flows. For the remaining OD pairs we assume zero-valued travel flows. An important aspect is the issue with how to deal with zero flows. A large fraction of zero-valued OD flows would definitely point towards a Poisson or a

(zero-inflated) negative Binomial interaction model. However, neither a Poisson nor a negative Binomial spatial autoregressive regression model for OD flows has been developed so far. We include income differences between Swiss communes as an explanatory variable in our models, since a higher income gives incentives to commute. In conclusion, we filter the initial flow matrix for inter-communal travel trips, income data available only in 1595 communes and all zero flows, which gives a final sample size of 46,659 OD flows. Clearly, this is a limitation of our modelling approach but nevertheless the findings can be of apparent value for pointing directions.

Modelling commuting behaviour requires a set of relevant explanatory variables that capture the characteristics of origins and destinations, along with the mechanisms that generate the trips among them. The dependent variable, inter-communal travel flows, is regressed on several independent variables obtained or derived from the 2000 Federal Census, the Swiss national transport model ARE (2005), and the Institute for Transport Planning and Systems (IVT) of ETH Zurich. We use the following variables in our framework, which are also common in explaining public transport demand in the literature (e.g. LeSage and Thomas-Agnan, 2015; Farmer, 2011; Axhausen et al., 2015).

Table 1: Model variables

Statistic	**Definition**
Flow	Av. daily flows
Network dist.	minutes
Income diff. rel	CHF (in 1,000)
Population (o)	# inhabitants
Jobs (d)	# jobs
Pop. density (d)	# pop. / area (in km^2)
Job density (o)	# jobs / area (in km^2)
Pop. accessibility (d)	# accessible pop.
Job accessibility (o)	# accessible jobs
Car (o)	# cars / pop.
Car (d)	# cars / pop.
Jobs3rd (o)	# jobs in 3rd sector / # jobs
Jobs3rd (d)	# jobs in 3rd sector / # jobs
Workers (o)	# workers3 / pop.
Workers (d)	# workers3 / pop.
Note:	(o),(d) = at origin, at destination municipalities

The starting point is a log form least-squares gravity model for OD flows in the form of

$$\log(y) = \alpha \log(l_N) + \beta_o \log(X_o) + \beta_d \log(X_d) + \delta \left(\frac{inc_d - inc_o}{inc_o}\right) + \gamma \log(g) + \epsilon,$$

where X_o and X_d are characteristics of origins and destinations, g denotes the network distance and $((inc_d - inc_o)/inc_o)$ reflects the relative difference of income between destination and origin municipalities. Estimation results of the gravity model are showed in Table 2. All parameters are highly significant except those of the share of 3rd sector jobs at origins and the share of cars per origin municipality having p-values lower than 5% and 1% respectively. The network distance decay parameter (-1.537) is within the expected range for commuting patterns, in accordance with previous studies. All other explanatory variables have a much weaker impact on the dependent variable, but this finding is in line with the expectations of existing literature (LeSage and Thomas-Agnan, 2015; Farmer, 2011). Income differences between destinations and origins have a significant and positive effect on travel-to-work trips and should be interpreted as an elasticity, since relative differences are used. This intuitively makes sense. By applying Moran's I tests we find that the residuals of the aspatial model indeed exhibit remaining spatial dependence and thus justify the need for spatial models. Spatial autoregressive models (SAR) are typically written as

$$y = \alpha\, l_N + \rho_i\, W_i\, y + \beta_o X_o + \beta_d X_d + \epsilon, \quad with\ i = o, d.$$

where the weights for W_i are defined as:

$$Netw.dist.weights: \quad w_{ij} = \frac{1}{traveltime_{ij}}, \quad Econ.dist.weights: \quad w_{ij} = \left(\frac{traveltime_{ij}}{\exp(((inc_d - inc_o)/inc_o))}\right)^{-1}$$

As it can be seen in Table 2, SAR models relying on origin- and destination-centric network and economic distance weights show positive influence of neighbouring communes on travel-to-work trips. Rho is higher than 1, which is an artefact of using the min-max approach for the spatial weights when building W_i, i = (o,d) instead of classic row-normalization (Kelejian and Prucha, 2010). In the transition from the gravity model to the SAR models, variables Car (o), Car (d), and Jobs3rd (o) are not statistically significant anymore and the impact of network distancebecomes smaller. Interestingly, rho for the SAR model relying on economic distance weights has a bigger impact compared to the network distance weighted SAR. It has to be emphasized that parameter estimates of spatial autoregressive regression models can not be interpreted as simple elasticities as in the gravity model, since spatial spillovers complicate the task of interpreting estimates from these models in a direct way. Furthermore, the spatial models yield a higher goodness-of-fit measure than the gravity model. Finally, we employ an IV regression framework to test the endogeneity of income and using a set of valid instruments (in line with Sarlas and Axhausen,2017). Note that pseudo R^2 values must be treated with caution, as they are not equivalent to OLS-based R^2 measures.

Table 2: Gravity model and spatial autoregressive models estimates

	Dependent variable: log(commuting flows)					
	Gravity model (OLS)		**SAR (GMM / 2IV)**		**SAR (GMM / 2IV)**	
			Network distance weights		*Econ. distance weights*	
	Estimate	Sign.	Estimate	Sign.	Estimate	Sign.
(Intercept)	4.443	***	5.765	***	5.788	***
log(Netw. distance)	-1.537	***	-1.250	***	-1.254	***
Rel. Income diff.	0.085	***	0.047	**	0.041	**
log(Jobs) (d)	0.473	***	0.307	***	0.308	***
log(Pop. density) (d)	0.030	***	0.036	***	0.036	***
log(Pop. access.) (d)	-0.176	***	-0.207	***	-0.206	***
log(Jobs3rd) (d)	0.102	***	0.082	***	0.082	***
log(Car) (d)	-0.071	***	-0.007		-0.006	
log(Workers) (d)	0.665	***	0.409	***	0.417	***
log(Population) (o)	0.440	***	0.359	***	0.358	***
log(Job density) (o)	-0.043	***	-0.042	***	-0.042	***
log(Job access.) (o)	-0.180	***	-0.239	***	-0.239	***
log(Jobs3rd) (o)	-0.027	**	-0.019		-0.018	
log(Car) (o)	-0.023	*	-0.003		-0.002	
log(Workers) (o)	0.365	***	0.249	***	0.254	***
rho			2.387	***	2.704	***
R^2	0.5177					
Pseudo adj. R^2			0.5898		0.5893	
HC robust std. errors	yes		yes		yes	

Note:

The resulting in-sample predictive accuracies outperform those from the current NPVM for different accuracy measures, as can be seen in Table 3.

Table 3: In-sample predictive accuracy measures

	RMSPE	RMdSPE	MAPE	MdAPE	SMAPE	SMdAPE
NPVM	87.50	6.49	271.76	64.92	74.12	64.24
Gravity model (OLS)	131.20	68.76	913.60	687.62	143.36	154.94
SAR (GMM / 2IV); Netw. dist. weigths	6.81	4.75	51.53	47.52	62.59	55.22
SAR (GMM / 2IV); Econ. dist. weigths	6.40	5.13	51.99	51.34	68.96	63.03