Spatial modelling of traffic volumes and mean speed values

Georgios Sarlas* & Kay W. Axhausen
Institute for Transport Planning and Systems (IVT), Swiss Federal Institute of Technology (ETH), Zurich, Switzerland

*Email: georgios.sarlas@ivt.baug.ethz.ch

1 Introduction

Econometrics have found application in a plethora of transport problems, constituting a medium both for obtaining parameters values, and for prediction purposes. However, applications acknowledging and treating for endogeneity and interdependence issues are rather sporadic (e.g. [1]). Furthermore, the spatial nature of the employed data makes spatial autocorrelation issues in many instances very likely [2]. The aforementioned issues, if overlooked, can lead to inconsistent and biased estimates, and even mis-specified models. Authors previous work dealt with the problem of estimating traffic volume and mean speed values independently of each other ([3],[4]), highlighting the strong presence of spatial autocorrelation issues and concluding on the use of spatial autoregressive models (SAR) as a remedy. Building upon that, in the current conference contribution we center our focus on providing a methodological framework capable of making both speed and volume predictions for any location of a network, accounting both for underlying endogeneity and spatial autocorrelation issues. The developed framework is subsequently tested in a case study set-up.

2 Methodology

Conceptually, it can be argued that the interdependence between speed and volume is apparent. Looking closer at the mechanism governing their interaction, speed is essentially the outcome of the demand and supply interaction at a link level. The result of that interaction might as well have a spillback effect through propagation of the congestion to the preceding links. On the other hand, lower speed values have an impact on the travel costs and thus on demand. Therefore, it can be concluded that this mechanism should constitute a core element when trying to model those phenomena. In the context of simulation (e.g. four-step model, agent-based model), that interaction is achieved through an iterative way until an equilibrium point has been reached. On the other hand, in the context of statistical modelling, that mechanism needs to be typified in order to isolate the variables that exert impact on the dependent variables directly, and indirectly through the other variable.

On the estimation front, we are making use of the two-step generalized method of moments (GMM) and instrumental variable (IV) estimation approach (denoted as 2SGM/IV), as proposed in [5] and [6]. This estimation approach allows both for endogenous regressor(s), heteroscedastic disturbances, and treatment of spatial effects while the estimation process entails four steps. In summary, at the outset a two-stage least squares (2SLS) estimator is applied, followed by a GMM estimator, then by a generalized spatial 2SLS estimator, and last by a GMM estimator to obtain the consistent and efficient estimates. The endogenous regressor(s) can be "replaced" either by an instrument variable, or can be estimated based on a set of exogenous regressors. Making use of the same estimation process the problem can be easily formulated as a structural equations model as well [7]. The general formula of the SAR model allowing for both spatial lagged dependent variables and error spatial dependence is:

$$Y = \rho WY + \beta_i X_i + u, with\, u = \lambda Wu + \varepsilon$$

Where ρ is the spatial autocorrelation parameter, λ the spatial autoregressive coefficient, W the spatial weight matrix, Y the dependent variable, X the independent variables, u the disturbances and ε the error term.

3 Case study

Making use of the case study set-up as presented in [4] (briefly, the nationwide network of Switzerland with 416 count locations), we proceed to the estimation of the models. Using as a departure point the previously estimated SAR error model [4] for average daily traffic volume (AADT), we tested whether the inclusion of the speed variable in the specification is statistically significant, which was not the case. As mean speed values we took the estimated ones from the Swiss national transport model (NPVM), which will be updated with true values in the near future (Tom-Tom travel time estimates). Before proceeding it is useful to put into perspective the assumptions of that model. At first, that the network structure does not have an endogenous impact (meaning that for a given network we can apply the model to estimate the AADT). Nevertheless, speed is assumed to have an indirect impact on the volume through the route choice process employed for the construction of the so-called accessibility weighted centrality variable. However, that centrality variable was constructed as a proxy of interregional demand and hence it can be assumed that the influence of speed is negligible and not directly related with the traffic volume.

Thereupon, only one endogeneity is considered (volume on the speed model). Previous attempts to associate volume and speed values on a macroscopic level normally employ different BPR curves (e.g. [8]), including the volume and the capacity of the links. Plotting the various mean speed values versus the capacity-volume ratio for our case (Fig.1), it appears that a function is present but it remains a task to specify its exact form and test for spatial autocorrelation issues.

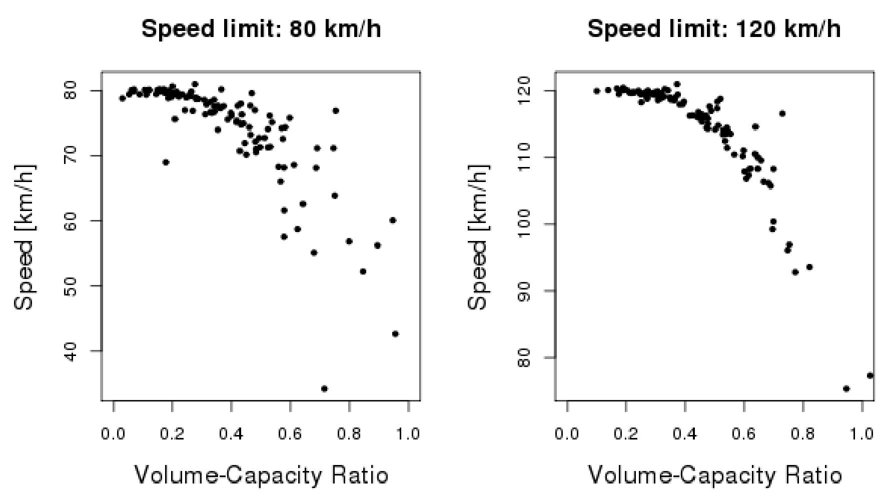

Figure 1: Mean speed vs volume-capacity ratio plots for two speed limit categories

To facilitate the estimation procedure, the following general form of travel time per link is defined. It should be noted that as Δtt we define the difference between the mean travel time and the free-flow travel time (based on the posted speed limits):

$$tt = to\left(1 + \frac{AADT^{\beta}}{Capacity^{\gamma}}\right)$$

$$tt - t_0 = {t_0}^{\alpha} * \frac{AADT^{\beta}}{Capacity^{\gamma}}$$

$$\log(\Delta tt) = \alpha * \log(t_0): link_{type} + \beta * \log(AADT) + \ \delta * \log(Capacity) + \varepsilon, with\ \delta = -\gamma$$

The model is estimated making use of the 2SGM/IV estimator, instrumenting the endogenous variable of volume with predictions made from a SAR volume model, which is estimated in terms of a GMM estimator. This way we allow for the treatment of spatial effects and heteroscedasticity on both models, and at the same time we account for the endogenous relationship between speed and volume.

4 Results

The model estimates are provided in the following table, along with the spatial autoregressive coefficient. Initially, the full spatial model was calculated, where the results supported the choice of a spatial model with error spatial dependence, which was also the case for the AADT model. This finding surfaces that spatial autocorrelation arises either due to missing variables, or due to shocks which cannot be captured in the model formulation. Especially, the latter is consistent with the kinematic wave theory that governs such fundamental relationships. A particular attention was given to the identification of the neighborhood extent (*W*), testing thoroughly both Euclidean and network distance-based weighting schemes. We conclude on the use of a network distance-based *W.*

Table 1: Model estimation results

Depend: variable : log *(ΔTT)*	Estimate	Std. Error	t-value	Sign.
log(capacity)	-3.82	0.23	-16.81	***
log(free-flow travel time:) Freeway	1.19	0.13	9.38	***
log(free-flow travel time): Major	0.80	0.17	4.80	***
log(free-flow travel time:) Urban main	0.27	0.75	0.36	
log(free-flow travel time): Rural major	1.68	0.20	8.31	***
$\log(\widetilde{AADT})$: Freeway	3.60	0.24	14.95	***
$\log(\widetilde{AADT})$: Major	3.97	0.22	17.84	***
$\log(\widetilde{AADT})$: Urban $main$	4.05	0.42	9.57	***
$\log(\widetilde{AADT})$: Rural major	3.44	0.24	14.54	***
Rho	0.86	0.40	2.14	*
*# of observations: 416; Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*				

Last, the results will be updated by taking into account actual mean speed observations which will allows us to draw solid conclusions with respect to the predictive accuracy of such a model and also in comparison with the NPVM.

References

[1] V. Shankar and F. Mannering, “Modeling the endogeneity of lane-mean speeds and lane-speed deviations: a structural equations approach”, *Transportation Research Part A: Policy and Practice*, vol. 32, no. 5, pp. 311–322, 1998.

[2] L. Anselin, *Spatial Econometrics: Methods and Models*, vol. 4. Dordrecht: Springer Netherlands, 1988.

[3] G. Sarlas and K. W. Axhausen, “Localized speed prediction with the use of spatial simultaneous autoregressive models”, p*aper presented at 94th Annual Transp. Res. Board Meeting, Washington D.C.*, 2015.

[4] G. Sarlas and K. W. Axhausen, “Prediction of AADT on a nationwide network based on an accessibility- weighted centrality measure”, *Arbeitsberichte Verkehrs- und Raumplanung,* no. 1094, IVT, ETH Zurich, Zurich, 2015.

[5] H. H. Kelejian and I. R. Prucha, “A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model,” *Int. Econ. Rev. (Philadelphia).*, vol. 40, no. 2, pp. 509–533, 1999.

[6] D. M. Drukker, P. Egger, and I. R. Prucha, “On Two-Step Estimation of a Spatial Autoregressive Model with Autoregressive Disturbances and Endogenous Regressors,” *Econom. Rev.*, vol. 32, no. 5–6, pp. 686–733, 2012.

[7] H. H. Kelejian and I. R. Prucha, “Estimation of simultaneous systems of spatially interrelated cross sectional equations,” *J. Econom.*, vol. 118, no. 1–2, pp. 27–50, 2004.

[8] National Research Council (US), “Metropolitan Travel Forecasting: Current Practice and Future Direction”, *Transportation Research Board*, vol. 288, 2007.