Quantifying the agglomeration effects of Swiss public transport between 2000 and 2010

Georgios Sarlas, IVT, ETH – Zürich
Raphael Fuhrer, IVT, ETH – Zürich
Kay W. Axhausen, IVT, ETH – Zürich

Conference paper STRC 2015

STRC | 15th Swiss Transport Research Conference
Monte Verità / Ascona, April 15-17, 2015

Quantifying the agglomeration effects of Swiss public transport between 2000 and 2010

G.Sarlas*, R. Fuhrer, K.W. Axhausen

Institute for Transport Planning and Systems (IVT), ETH Zurich, Stefano-Franscini-Platz 5, CH 8093, Zurich, Switzerland

Abstract

This conference contribution deals with the economic effects of public transport supply improvements on productivity, generally referred to as agglomeration effects. To the knowledge of the authors, there has not been a systematic study regarding these effects in Switzerland. To enable that, the development of productivity on a municipal level between 2000 and 2010 is regressed on variables, such as private and public transport accessibility, and variables describing the local economic structure. Productivity is operationalized by the average amount of salaries paid at the location of their generation. Different regression models are tested to account for the spatial and temporal character of the data. We conclude that agglomeration effects are present, quantifiable, and for the case of public transport supply they are found to take values between 1 and 3% for the nationwide spatial analysis level, while for the agglomeration areas the effects are significantly higher and take values between 2 and 4%. Moreover, elasticity values exhibit variation over the analysis period reflecting a differentiated dynamic of the externalities that can result due to the public transport improvements.

Keywords

Agglomeration effects, public transport improvement, externalities, spatial econometrics, geographically weighted regression

**corresponding author*
Email address: georgios.sarlas@ivt.baug.ethz.ch
Telephone: +41-44-633 37 93

1. Introduction

“Throughout the evolution of human settlements, there is only one factor which defines their extent: the distance man wants to go or can go in the course of his daily life. The shortest of the two distances defines the extent of the real human settlement, through definition of a daily urban system" (Doxiadis, 1970). Naturally, the importance of the transportation aspect comes to the surface in defining the spatial extent of human settlements, which differs from the prevailing perception of the physical structure ("the built-up area is the city") or the institutional frame ("the municipality is the city") of the human settlements. In order to overcome this limitation, the term urban agglomeration was adopted to specify the extent of human settlements around main cities where major economic activities are concentrated. As a result of this spatial concentration of economic activity, externalities arise that can lead to positive effects on productivity (Graham, 2007).

Duranton and Puga (2004), in their discussion of micro-foundations of urban agglomeration economies, mention three main mechanisms that are responsible for agglomeration economy effects: sharing, matching and learning mechanisms. By sharing, the mechanism that allows firms and individuals to share the same input (e.g. transport network) can lead to a wider economic gain. By matching, the mechanism that allows the firms and the labour to have a better matching on their requirements and needs (e.g. specialized labour), and thus facilitate a better matching between them, leading in turn to an increase in the productivity. By learning, the mechanism that allows the knowledge transfer to happen and lead to increased productivity.

As Chatman and Noland (2011) argue, public transport improvements are capable of having substantial external benefits by enabling agglomeration economies. This is facilitated by increasing the accessibility between the firms and also among firms and labour force as the result of improved transport connection, and thus reduced travel cost. Lowering generalized transport costs realizes increased accessibility between firms and labour force. This increases the chances that a company is able to find exactly the employee it is looking for, or that a person is able to find a job position that exactly matches her or his qualification and talents. In this situation, specialization of firms but also of persons is enhanced, both involved parties are better off and that subsequently leads to increased productivity. In the same way, agglomeration benefits result by the creation of new jobs, since an increase in the employment density results to increased job opportunities for the labour at the same cost as before, and thus gives rise to production gains.

In the case of Switzerland, a state owned dense public transport network (rail, roads, rivers and lakes) exists. In the 1980s, Swiss people voted for a huge rail improvement plan, including network extensions, network and stations capacity increase and new fleet. As a result, an interval timetable was established soon and many lines were operated quicker, more frequent and noticeably coordinated. Notably, the biggest change at one-go happened in 2006 when a new transalpine tunnel and a high-speed track on key locations of the network were opened. Overall, public transport has become much more attractive due to improvements in the level of service and travel time savings due to frequent headways, coordinated connections and optimized lines.

To the knowledge of the authors, there has not been a systematic study regarding the agglomeration effects in Switzerland, and in particular of the benefits that accrue as the result of the improvement of the public transport supply. Therefore, the current study aims to fill in that gap and investigate the existence of such benefits and attempt to quantify them.

2. Methodology

In order to facilitate the identification and the quantification of these benefits to take place, regression analysis is the most appropriate method for such purposes. More specifically, regression analysis constitutes a statistical process that estimates the magnitude of the relationship between a set of independent variables and the dependent variable of interest, hence quantifying the underlying causality, if any.

A key aspect on the determination of the methodology is how to put into operation the growth and the public transport network, under the context of regression analysis. At first, economic growth can be measured in terms of productivity. Several direct and indirect methods of capitalizing productivity can be found in the literature, each one associated with different difficulties in terms of data demand. As Chatman and Noland (2011) mention, salaries and gross domestic product constitute the most typical measures of productivity, while more sophisticated methods, looking at the contribution of labour and capital to firm revenues, exist. In the case of salaries, the underlying assumption is that as the productivity increases, this leads in turn to raises to the employees' salaries. Thereupon, the economic index of the generated salaries is considered as an adequate approximation to capture changes in productivity over time.

The next critical aspect of the methodology is how to quantify the public transport supply in a way that allows the improvements over time to be communicated in the modelling approach. In a similar line of thought with the definition of the effective density of employment by Graham (2007), the measure of travel accessibility is employed, which is a measure of how far people are willing, or able, to travel on the course of their daily life and quantifies how interaction opportunities decrease over the distance. Since Hansen (1959) first formulated accessibility in mathematical terms, a variety of different approaches have evolved (e.g. see Kwan, 1998). All of those approaches have in common the following; first they heavily rely on a distance decay function, second on a transport infrastructure model and third on spatial densities of so-called activity or opportunity points (e.g. inhabitants, work places). A gravity-like formulation is used:

$$A_i = \Sigma O_j * e^{\beta * C_{ij}} \quad (1)$$

A_i being the accessibility in point *i* to all *j* opportunity points O_j at generalized costs C_{ij} that are weighted by a negative exponential transformation with a factor β. Apparently, factor β is crucial for the calculation of the accessibility values which captures how far people want to travel, and it is thus associated with the behavioural dimension of travellers, regarding their trip length. Naturally, different accessibilities (and thus beta parameters) are associated with different trip purposes and modes; people are willing to travel shorter distances for shopping activities than for e.g. commuting to work. The inclusion of the generalized costs in the accessibility formulation is chosen to replace the distance metric by a cost metric which is more representative since it incorporates information about the network connectivity and also about people's perception of travel time.

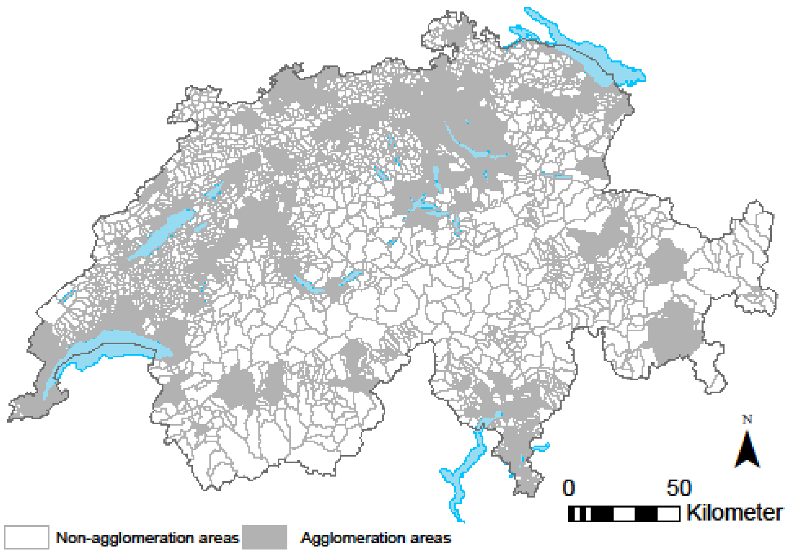
Productivity per municipality is modelled as a function of variables. In particular, the set of variables includes two wider components controlling for the impacts on productivity; the first one describes the socio-demographic characteristics of the labour force (e.g. age, education, qualifications, gender, type of work etc.), while the second one describes the spatial variation of the characteristics of the municipality where the economic activity takes place (e.g. economic structure, transport accessibility etc.). In summary, the inclusion of two variables capturing agglomeration benefits is of interest and needs to be tested; first the impact of the spatially concentrated economic activity (employment density), and second the labour market accessibility which allows a better match between an employer's needs and employee's skills. One the main challenge is the isolation of the transport effects from other possible sources of productivity gains (Graham and Van Dender, 2011), as well as the simultaneity in transport supply and productivity. It becomes apparent, that the successful isolation of the public transport effects from the private transport ones, constitutes one of the main issues that the current research deals with.

A note should be made on the assumed causality aspects of the models. More specifically, endogeneity issues exist between the productivity and the agglomeration effects of transport and the employment density, since the improved supply of transportation might as well be the result of increased productivity (through more generated taxes to finance transport related investments), and increased density might be the result of increased productivity (as increased attraction). Nevertheless, that constitutes a limitation that is not addressed. However, as Melo et al. (2009) argue, applying corrections for accounting for the reverse causality of agglomeration issues do not appear to produce noticeable changes in the estimates.

Different formulations of models are developed and tested on their capability of quantifying adequately the agglomeration effects, within the framework of a production function, where the observed variation of the employed economic indices across the years (2000,2005 and 2010) is related to changes in the accessibility. In particular, three different levels of regression are employed. At first, the OLS model where a different model per year is estimated under the assumption that the error terms of the model are independent and identically distributed otherwise it can give rise to biased estimated coefficients. Secondly, the temporal correlation of the observations is taken into account in the formulation of the model through the estimation of panel data models. Last, an important aspect that is taken into account is the implications of using spatial data in the models. More specifically, the use of such data might lead to the existence of spatial dependence, which in turn can lead to biased estimations and therefore needs to be accounted in properly in the model formulation through applying the proper corrections (family of spatial simultaneous autoregressive models (SAR), an overview can be found at LeSage and Pace, 2004), while the issue of spatial heterogeneity that governs economic activities (whether a structural equation holds over space) is investigated through the use of geographically weighted regression (GWR) that provides localized coefficients (an overview can be found at Charlton and Fotheringham, 2009). The spatial dependence of data is accounted for in the panel data formulation as well, while that is not feasible for the GWR formulation where an approved methodology for estimating of GWR panel data models is still not present in the literature.

In addition to the different regression models, two spatial levels of analysis are employed to quantify the agglomeration effects and also to control for the impact of urbanized areas on them. More specifically, one level of analysis corresponds to the whole country and aims to provide nationwide estimates, while a second one corresponds to the agglomeration areas only, as those are defined by BfS (see Figure 1), and aims to provide estimates that concern only the urbanized areas with notably higher concentration of economic activities, and consequently higher relevant contribution to the nationwide productivity. In summary, all the estimated models aim to provide the rate of how much salaries change when private transport and public transport accessibility change respectively (elasticity), which is the relationship of interest to check its existence and quantify it. In addition, the elasticity of local employment density is quantified as well which constitutes also an agglomeration effect.

Figure 1: Geographical definition of the agglomeration areas



Data: Agglomeration and metropolitan areas, BfS 2000

3. Data

In order to facilitate an empirical analysis on the existence of agglomeration effects of public transport to take place, various data need to be utilized. In particular and in accordance to the proposed methodology, three sources of data are required. First, data capable of capitalizing the development of productivity is required. Productivity is approximated by the average amount of salaries paid at the location of their generation, utilizing data coming from the BfS salary data survey (Lohnstrukturerhebung), collected every two years. Second, data with respect to the transportation is needed for the calculation of the travel accessibility values. The national transport models of years 2000,2005 and 2010 (ARE, Nationales Personenverkehrsmodell) provide the required data for that purpose. Last, spatial data on the population and the employment per municipality are needed for the calculation of the accessibility values and also for the determination of the spatially concentrated economic activity. These data are available from the BfS employment and population census accordingly, which they also exist in the structural data of the national transport model.

An important issue on the data front, is to bring all the data on the same geographical and temporal level, taking into account the various limitations that this entails. Salary data is reported per postcode area, while in the national model, a traffic zonal level, corresponding to the municipal level and the district level in the case of the main cities for the year 2000, is kept constant across all yearly versions of the national model. The municipal changes that have occurred since 2000 (mainly in the form of merges) are taken into account in the processing of the spatial data. In summary, the traffic zonal level, included in the national models is chosen to be the analysis level for the current study, including 2949 zones. On the temporal level front, the levels of the national model are chosen as the temporal analysis levels. This choice of temporal analysis level is partially consistent with the one of the salary data where the data are available for the years 2000, 2006 and 2010 respectively. Naturally to overcome this limitation, an assumption is made that the salary data of year 2006 are taken as approximation of year 2005. Data processing and model estimation are undertaken with the statistical programming language R (R Development Core Team, 2011).

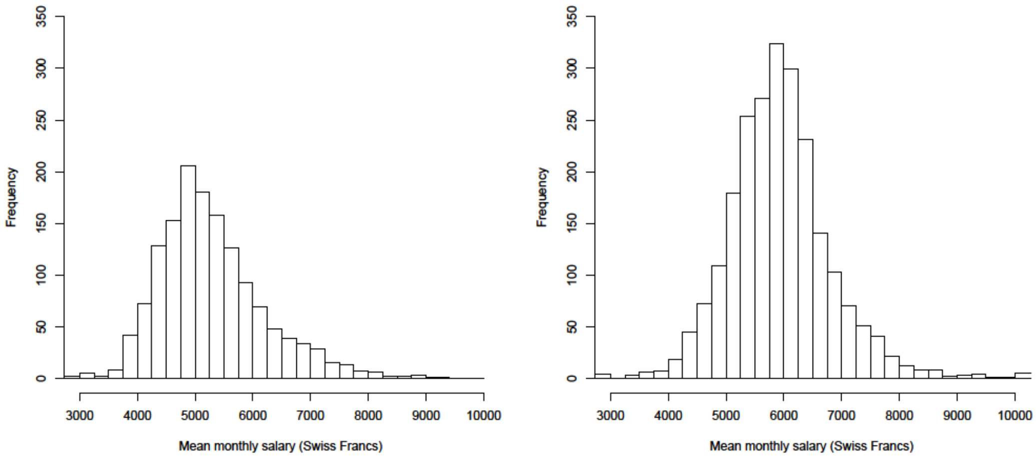
3.1 Salary data

Salary data include individual salary observations along with additional information regarding the qualification and the characteristics of the employees. More specifically, information on the qualifications, the education attainment, the age, the in-post duration, the management duties, the type of industry, are included among others in the salary data survey. Data is reported anonymously and it is not possible to track persons over time (unique id every year).

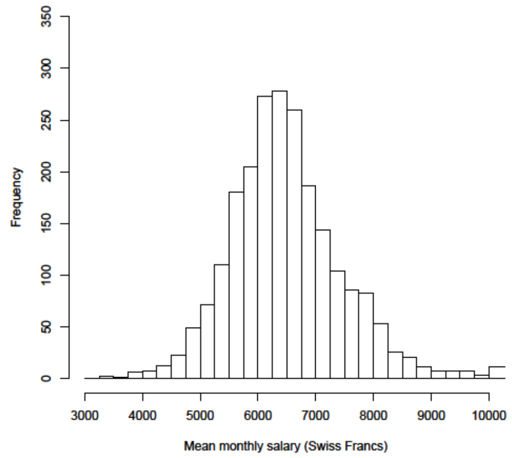
The information about where the salary is generated is included by a reported postcode, constituting a finer level than the employed analysis level. The matching of the postcodes to the employed spatial analysis level is made taking into account the municipal changes over years and demanded utilization of freely available data from the Swiss post-office, geographical processing and manual matching, especially in the cases of the main cities where a finer level is adopted than the data from the post-office.

The individual salary observations are firstly aggregated per postcode, and subsequently per traffic zone to construct the variables that are later used in the regression models. In summary, for each zone, based on the sample of the individual observations, the average salary is taken as representative, while the corresponding percentages of people’s qualifications, education attainment etc., are calculated in a similar way and assumed to be representative of the economic structure of each zone. In Figure 2, the histograms of the average salary per zone are presented. It should be noted that in the year 2000 we have the least spatial coverage of the analysis level in terms of availability of salary data per zone (1600 out of 2949 zones), while in the following years the coverage is significantly higher (approximately 2300 zones).

Figure 2: Histograms of average salary per zone



(a) Year 2000 (b) Year 2005



(c) Year 2010

3.2 National transport model

The national transport models include information about how and where people are travelling to, corresponding to the commuting trips for working purposes. In particular, the corresponding origin-destination matrices along with the generalized cost matrices for car and public transport respectively, are of interest for the present work. In the case of car's generalized cost, this is constructed by taking into account only the in-vehicle time, while for the public transport the access/ egress time, the waiting time, the number of transfers and the in-vehicle.

Moreover, apart from the transport-related information extracted from the national model, the available structural data incorporated into it are utilized. That data includes information on the population and the employment positions per zone. In order to obtain more detailed information on the economic structure of the zones (3^{rd} sector percentage), the employment survey data is used to accompany the existing data of the national model.

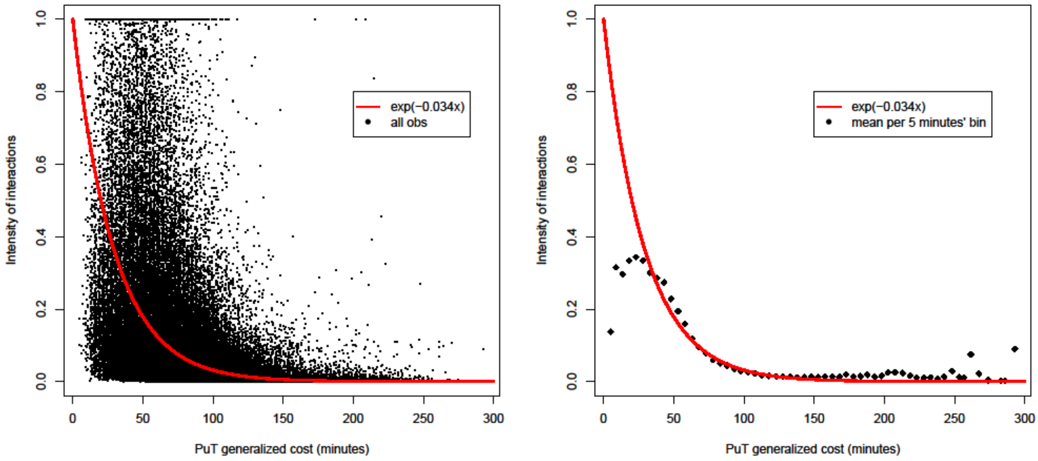
3.3 Travel accessibility

Having obtained the required data for the formulation of the accessibility measures, the next critical point is the definition of the beta parameter. Applying the estimated betas from previous studies for the case of Switzerland has a number of drawbacks, since they are bound heavily to the datasets used for their estimation. In order to overcome this impediment, the accessibility parameters are estimated from the available data of the national model. The estimation takes place following a similar methodology as presented by Halás et al. (2014) for the estimation of the accessibility parameter of urban centres. However, in the present study a global measure of accessibility is used instead, given the country-level productivity that is of interest. The estimation of the betas takes place by defining the portion of daily commuters, from each traffic zone to each other zone, out of the total out-commuters of the origin zone. These portions take values between 0 and 1 and they are referred to as the interaction intensity. A normalization of the aforementioned portions of each zone by the according maximum percentage follows to ensure that we have values covering the whole range of potential values. Subsequently, the next step is to quantify how interaction intensity decreases over space, which actually corresponds to the beta parameter of the accessibility formulation. For each O-D pair with a value of interaction intensity higher than zero, we extract the corresponding generalized cost per mode and construct the datasets for the estimation of the beta parameter per year and mode, according to the accessibility formulation. The nonlinear least-squares estimates of the beta parameters happen by following Gauss-Newton algorithm. The resulted beta parameters are summarized in Table 1, while in Figure 3 the fit of the estimated betas for year 2010 is presented, both for disaggregated and aggregated data per 5 minutes' time intervals. Interestingly, betas for car trips result to a much steeper curve than for public transport, indicating that people are more willing to bear higher generalized costs for travelling by public transport. However, attempting a direct comparison of the different betas per mode should be done with caution and definitely by taking into account that the constructed generalized costs per mode differ significantly, to avoid drawing wrong conclusions.

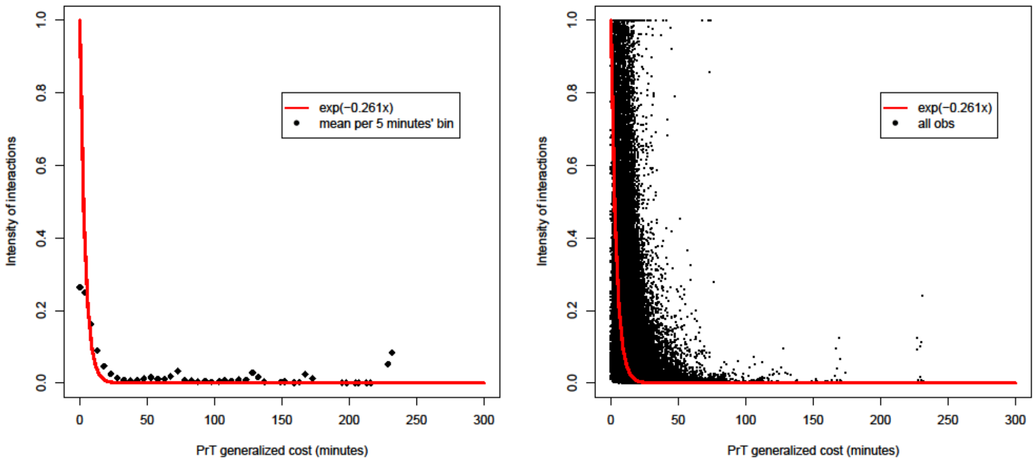
Table 1: Estimated accessibility parameter per mode and year

Mode	2000	2005	2010
Public transport	-0.0312	-0.0323	-0.0344
Car	-0.2960	-0.2950	-0.2613

Figure 3: Estimated beta parameters for year 2010



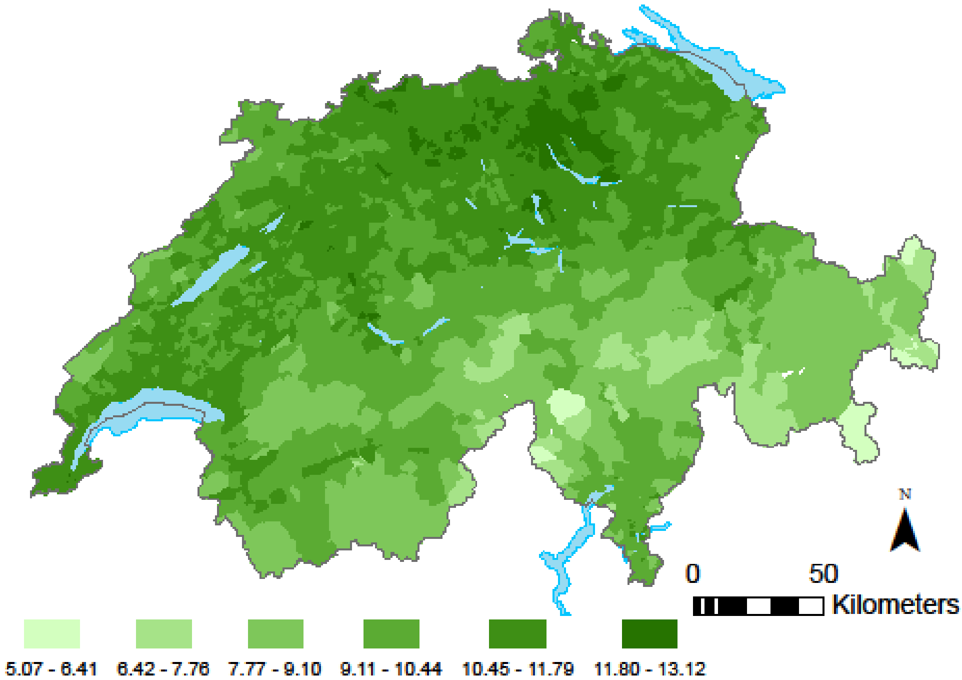
(a) Public transport accessibility parameter



(b) Car accessibility parameter

Having estimated the beta parameters of the accessibility, the next step is to proceed to the calculation of the population accessibility values per mode and year. The logarithmic values of the accessibilities are of interest and are included in the model specification, as a measure of the availability of labour pool per zone. Public transport accessibility values in year 2010 are visualized in Figure 4 to put into perspective its spatial variation. More specifically, higher accessibility values exist around the main cities of the country while the southern and southern-east part of Switzerland have the lower values, especially in the rural areas.

Figure 4: Calculated population log accessibility values for public transport in year 2010



Data: National transport models 2010 and BfS

4. Results

In this section different model formulations are developed and tested to assess their capability of quantifying adequately the agglomeration effects. The dependent variable is chosen to be the logarithm of the average salary per municipality, as an approximation of productivity, while accessibilities and employment are included in the model formulation in a logarithmic form as well. This format of variables provides directly elasticity values which are the measures of interest for the current study. At first, in the table below the summary statistics of the aggregated dataset used for the estimation of the models are put into perspective.

Table 2: Summary statistics of dataset (all years)

Variable	Min.	1st Quart.	Median	Mean	3rd Quart.	Max.	Unit
Ln mean salary	7.86	8.58	8.69	8.68	8.79	9.67	Log
Ln car accessibility	1.78	8.81	9.45	9.25	9.93	12.01	Log
Ln public transport accessibility	5.51	10.07	10.70	10.56	11.21	13.13	Log
Ln number of local employed (FTE)	1.10	5.34	6.38	6.42	7.33	11.00	Log
Commuter from outside Switzerland	0.00	0.00	0.00	0.05	0.05	0.91	%
Short residence permit	0.00	0.00	0.00	0.02	0.01	0.73	%
Average duration in-post	0.16	7.69	9.19	9.40	10.80	23.60	Years
Ln average age	3.13	3.69	3.73	3.72	3.76	4.05	Log
Men	0.00	0.46	0.56	0.56	0.67	1.00	%
Tertiary education	0.00	0.00	0.01	0.04	0.05	0.82	%
Professional training	0.00	0.00	0.03	0.05	0.06	1.00	%
Further vocational training	0.00	0.04	0.07	0.08	0.11	1.00	%
Teaching degree	0.00	0.00	0.00	0.03	0.02	1.00	%
Highschool diploma	0.00	0.00	0.00	0.01	0.02	0.42	%
Vocational training	0.00	0.40	0.53	0.52	0.63	1.00	%
Positions with highest demands	0.00	0.02	0.05	0.05	0.07	0.92	%
Positions with qualified indep. work	0.00	0.15	0.24	0.26	0.33	1.00	%
Positions with professional skills	0.00	0.33	0.41	0.42	0.50	1.00	%
Working (other private sector)	0.00	0.68	0.85	0.77	0.94	1.00	%
Working (manufacturing)	0.00	0.00	0.00	0.03	0.01	1.00	%
Working (FIRE)	0.00	0.00	0.01	0.03	0.03	1.00	%
Working (hotel, restaurants)	0.00	0.00	0.00	0.04	0.02	1.00	%

4.1 OLS regression

The first category of models is the OLS model for each year separately. The inclusion of different variables is conducted on the basis of their capability to improve the predictive accuracy of the model, and in particular by taking into account the adjusted R square as a measure of fit (describes the explained variance of the data), and their statistical significance. The specification of the model is kept constant over the years on purpose, to exhibit adequately the temporal variance of the estimated coefficients both in terms of values, but also in terms of statistical significance.

The estimated models for the two spatial analysis levels are presented in the following tables (Table 3and Table 4). Furthermore, the residuals of both regressions are tested for heteroscedasticity following Kroenker's test (Kroenker, 1981), which is a studentized version of the traditional Breusch and Pagan test (Breusch and Pagan, 1979) and shows that in both cases residuals are highly heteroscedastic. Corrections are made by using the HC0 estimator (White, 1980), resulting to heteroskedasticity-consistent estimation of the covariance matrix of the coefficient estimates in the regression models, and thus obtain the correct p-values for the statistical significance of the estimates.

Table 3: OLS Regression

Independent variable: Ln mean salary	Year 2000		Year 2005		Year 2010	
	Estimate	Pr(>|t|)	Estimate	Pr(>|t|)	Estimate	Pr(>|t|)
Intercept	6.538	***	6.923	***	6.864	***
Ln car accessibility	0.010	**	0.017	***	0.011	**
Ln public transport accessibility	0.018	***	0.016	***	0.015	***
Ln number of local employed	0.016	***	0.012	***	0.014	***
Commuters from outside Switzerland	-0.121	***	-0.087	***	-0.097	**
Short residence permit	-0.189		-0.147	*	-0.189	.
Average duration in-post	0.003	*	0.007	***	0.005	***
Ln average age	0.336	***	0.267	***	0.319	***
Men	0.176	***	0.059	**	0.126	***
Tertiary education	0.900	***	0.691	***	0.594	***
Professional training	0.520	***	0.22	***	0.317	***
Further vocational training	0.210	***	0.187	***	0.233	***
Teaching degree	0.169	*	0.192	***	0.321	***
Highschool diploma	0.620	***	0.236	*	0.253	.
Vocational training	0.063	**	0.035	.	0.020	
Positions with highest demands	0.436	**	0.408	***	0.397	***
Positions with qualified indep. work	0.203	***	0.255	***	0.242	***
Positions with professional skills	0.142	***	0.200	***	0.145	***
Working (3rd sector)	0.200	**	0.157	***	0.071	
Working (private sector)	-0.116	***	-0.109	***	-0.071	***
Working (manufacturing)	-0.222	***	-0.245	***	-0.102	**
Working (FIRE)	0.144	***	-0.01		0.061	
Working (hotel, restaurants)	-0.140	***	-0.127	***	-0.108	***
Residual Standard Error		0.095		0.087		0.095
Adjusted R-squared		0.676		0.647		0.603
AIC		-2675		-4651		-4143
# observations		1448		2298		2229

*Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

Table 4: OLS regression for the agglomeration areas

Independent variable: Ln mean salary	Year 2000		Year 2005		Year 2010	
	Estimate	Pr(>|t|)	Estimate	Pr(>|t|)	Estimate	Pr(>|t|)
Intercept	6.453	***	6.252	***	5.870	***
Ln car accessibility	0.027	***	0.022	***	0.027	**
Ln public transport accessibility	0.021	**	0.039	***	0.027	***
Ln number of local employed	0.018	***	0.010	***	0.015	***
Commuters from outside Switzerland	-0.118	**	-0.084	*	-0.115	
Short residence permit	-0.075		0.216	.	0.514	
Average duration in-post	0.005	*	0.007	**	0.003	
Ln average age	0.278	*	0.316	***	0.445	***
Men	0.141	**	-0.015		0.054	
Tertiary education	0.898	***	0.707	***	0.625	***
Professional training	0.455	***	0.340	***	0.240	***
Further vocational training	0.250	**	0.209	**	0.247	***
Teaching degree	0.409	***	0.284	***	0.462	***
Highschool diploma	0.606	**	0.462	**	0.379	
Vocational training	0.067	*	0.104	**	0.056	
Positions with highest demands	0.284		0.486	**	0.644	***
Positions with qualified indep. work	0.304	***	0.214	***	0.290	***
Positions with professional skills	0.136	***	0.120	***	0.195	***
Working (3rd sector)	0.215	*	0.377	***	0.281	***
Working (private sector)	-0.064	**	-0.069	***	-0.046	
Working (manufacturing)	-0.289	***	-0.168	**	-0.013	*
Working (FIRE)	0.147	***	0.168	*	0.174	***
Working (hotel, restaurants)	-0.179	***	-0.106	*	-0.127	***
Residual Standard Error		0.088		0.084		0.09
Adjusted R-squared		0.734		0.683		0.669
AIC		-1576		-2001		-1861
# observations		789		958		954

*Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1*

As it can be seen in the OLS results, a significant variation of the estimated parameters through the years is observed. More specifically, it is worthwhile to highlight that the estimated coefficients of the public transport accessibility indicates a decreasing impact over the years, decreasing from 1.8% elasticity value to 1.5%, for the nationwide spatial analysis level. However, the pattern differs in the case of the agglomeration areas’ analysis level, where public transport accessibility is found to increase from 2.1% to 3.9%, and then decrease to 2.7%. Furthermore, it becomes apparent that the agglomeration effect of public transport is higher in the finer spatial analysis level, reflecting the higher dependence of urbanised areas on transport systems. In addition, it should be mentioned that in the year 2000 the dataset has the smallest size due to the smallest spatial coverage, in comparison to the data sets of years 2005 and 2010 respectively. The elasticity of the local employment density is found to take values in the range of 1.2% - 1.6% in the nationwide results, and 1% – 1.8% in the agglomeration areas’ results, exhibiting the same pattern over the years. It should be noted that all estimated models correspond to regressions without the use of any weights, constituting a more conservative approach than weighted regression, since the employment of inappropriate weights might result to unrealistic results, over/underestimating the causality aspects.

4.2 Panel data analysis

The next category of models takes into account the temporal dimension of the observations. More specifically, two different models are estimated. The first one, denoted as pooling OLS, constitutes a general case of panel data formulation where the temporal dimension of the observations is neglected and it actually corresponds to the OLS where the relationship is assumed to be constant across the years (constant coefficients). The second model is a fixed time-effects model that controls for the impact of time on the estimated coefficients and assumes constant coefficients over the years as well. The impact of serial correlation through the estimation of fixed-individual effect models is tested as well but it gives no significant individual effects results due to the limited points in time.

The formulation of fixed time-effects model with random effects is tested as well but the corresponding statistical tests reject the use of such models. In particular, two Lagrange multiplier tests are calculated which test for time-effects in the residuals of the pooling OLS model (see Gourieroux et. al, 1982, Breusch and Pagan, 1980,and Honda, 1985). In addition, an F-test for time-effects is calculated as well, comparing pooling versus time-effects model formulations. All three previous tests, exhibit clearly that time-effects model should be used instead of pooling OLS, while a Hausman test (Hausman, 1978) for fixed versus random-effects model shows that the later formulation is inconsistent and thus fixed time-effects models should be preferred. All tests are estimated for both levels of spatial analysis and the results in both cases show the superiority of the time-effects model for the particular case at hand, indicating that pooling OLS might give rise to inconsistent estimates. The results of the tests are presented in Table 5.

Table 5: Panel data model effects tests

	Nationwide dataset		Agglomeration dataset	
	Estimate	Pr(>|t|)	Estimate	Pr(>|t|)
LM test - time effects (Breusch-Pagan)	159	***	73	***
LM test - time effects (Honda); chi-square	25181	***	5398	***
F-test for time-effects	385	***	185	***
Hausman test for fixed versus random effects; chi-square	3		3	

The estimation of the models takes place only for a balanced dataset, that corresponds to all the municipal observations that are included in all 3 different years datasets. In addition to the above, it is chosen to focus to the formulation of panel data models for the period 2005 – 2010, neglecting year 2000, in order to utilize the significantly higher number of common observations at these two points of time. On the heteroscedasticity front, the same statistical test as before is used (Kroenker's test) and for the case of the pooling OLs model the HC0 estimator (White, 1980) is applied, while for the fixed time-effects models a similar estimator is applied which restricts the common variance within each group of time period (more information can be found at Greene 2003).

The results for the nationwide datasets are presented inTable 6 and Table 7, while for the agglomeration areas in Table 8 and Table 9 accordingly.

Table 6: Pooling OLS and time-effects model

Independent variable: Ln mean salary	Pooling OLS		Time-effects	
	Estimate	Pr(>|t|)	Estimate	Pr(>|t|)
Intercept	5.416	***	6.223	***
Year 2005 dummy (time-effect)			0.082	***
Year 2010 dummy (time-effect)			0.118	***
Ln car accessibility	0.013	***	0.010	***
Ln public transport accessibility	0.012	***	0.020	***
Ln number of local employed	0.018	***	0.016	***
Commuters from outside Switzerland	-0.065	**	-0.098	***
Short residence permit	-0.040		-0.117	**
Average duration in-post	0.001		0.003	***
Ln average age	0.658	***	0.404	***
Men	0.132	***	0.134	***
Tertiary education	0.785	***	0.778	***
Professional training	0.427	***	0.399	***
Further vocational training	0.297	***	0.222	***
Teaching degree	0.381	***	0.369	***
Highschool diploma	0.431	***	0.356	***
Vocational training	0.074	***	0.075	***
Positions with highest demands	0.628	***	0.447	***
Positions with qualified indep. work	0.377	***	0.248	***
Positions with professional skills	0.241	***	0.169	***
Working (3rd sector)	0.082	.	0.198	***
Working (private sector)	-0.065	***	-0.075	***
Working (manufacturing)	-0.172	***	-0.207	***
Working (FIRE)	0.028		0.145	***
Working (hotel, restaurants)	-0.135	***	-0.126	***
Adjusted R-squared		0.676		0.742
Panel observations			1374 (total 4122)	

*Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

Table 7: Pooling OLS and time-effects model for years 2005-2010

Independent variable: Ln mean salary	Pooling OLS		Time-effects	
	Estimate	Pr(>|t|)	Estimate	Pr(>|t|)
Intercept	6.628	***	6.811	***
Year 2010 dummy (time-effect)			0.040	***
Ln car accessibility	0.021	***	0.014	***
Ln public transport accessibility	0.007	.	0.015	***
Ln number of local employed	0.014	***	0.013	***
Commuters from outside Switzerland	-0.085	***	-0.088	***
Short residence permit	-0.194	***	-0.156	***
Average duration in-post	0.006	***	0.006	***
Ln average age	0.360	***	0.308	***
Men	0.090	***	0.098	***
Tertiary education	0.653	***	0.630	***
Professional training	0.304	***	0.287	***
Further vocational training	0.238	***	0.230	***
Teaching degree	0.321	***	0.294	***
Highschool diploma	0.338	***	0.327	***
Vocational training	0.049	**	0.038	**
Positions with highest demands	0.404	***	0.411	***
Positions with qualified indep. work	0.278	***	0.247	***
Positions with professional skills	0.186	***	0.171	***
Working (3rd sector)	0.107	**	0.098	***
Working (private sector)	-0.081	***	-0.089	***
Working (manufacturing)	-0.172	***	-0.166	***
Working (FIRE)	-0.005		0.016	
Working (hotel, restaurants)	-0.086	***	-0.101	***
Adjusted R-squared	0.621			0.635
Panel observations				2117 (total 4234)

*Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

Table 8: Pooling OLS and time-effects model for the agglomeration areas

Independent variable: Ln mean salary	Pooling OLS		Time-effects	
	Estimate	Pr(>|t|)	Estimate	Pr(>|t|)
Intercept	5.448	***	6.199	***
Year 2005 dummy (time-effect)			0.081	***
Year 2010 dummy (time-effect)			0.107	***
Ln car accessibility	0.029	***	0.026	***
Ln public transport accessibility	0.018	***	0.027	***
Ln number of local employed	0.016	***	0.014	***
Commuters from outside Switzerland	-0.101	***	-0.129	***
Short residence permit	0.175		0.120	
Average duration in-post	0.002		0.005	***
Ln average age	0.570	***	0.326	***
Men	0.082	*	0.068	***
Tertiary education	0.732	***	0.735	***
Professional training	0.364	***	0.349	***
Further vocational training	0.260	***	0.176	***
Teaching degree	0.465	***	0.458	***
Highschool diploma	0.605	***	0.458	***
Vocational training	0.078	***	0.086	***
Positions with highest demands	0.696	***	0.551	***
Positions with qualified indep. work	0.441	***	0.301	***
Positions with professional skills	0.247	***	0.171	***
Working (3rd sector)	0.130	*	0.280	***
Working (private sector)	-0.027	.	-0.032	**
Working (manufacturing)	-0.135	**	-0.200	***
Working (FIRE)	0.035		0.151	***
Working (hotel, restaurants)	-0.175	***	-0.164	***
Adjusted R-squared		0.735		0.770
Panel observations				763 (total 2289)

*Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1*

Table 9: Pooling OLS and time-effects model for years 2005-2010 for the agglomeration areas

Independent variable: Ln mean salary	Pooling OLS		Time-effects	
	Estimate	Pr(>|t|)	Estimate	Pr(>|t|)
Intercept	5.906	***	6.082	***
Year 2010 dummy (time-effect)			0.032	***
Ln car accessibility	0.033	***	0.024	***
Ln public transport accessibility	0.026	***	0.033	***
Ln number of local employed	0.013	***	0.013	***
Commuters from outside Switzerland	-0.100	***	-0.102	***
Short residence permit	0.167	.	0.246	**
Average duration in-post	0.005	*	0.005	***
Ln average age	0.422	***	0.375	***
Men	0.021		0.024	
Tertiary education	0.661	***	0.650	***
Professional training	0.281	***	0.279	***
Further vocational training	0.234	***	0.232	***
Teaching degree	0.426	***	0.411	***
Highschool diploma	0.354	*	0.366	***
Vocational training	0.064	*	0.068	**
Positions with highest demands	0.557	***	0.558	***
Positions with qualified indep. work	0.291	***	0.255	***
Positions with professional skills	0.176	***	0.155	***
Working (3rd sector)	0.302	***	0.307	***
Working (private sector)	-0.039	*	-0.047	***
Working (manufacturing)	-0.126	.	-0.111	***
Working (FIRE)	0.145	*	0.174	***
Working (hotel, restaurants)	-0.097	*	-0.110	***
Adjusted R-squared	0.674			0.681
Panel observations				930 (total 1860)

*Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

4.3 Spatial regression models

The next category of models is the spatial regression models. Two categories of spatial regression models are tested, each one serving different purpose and having different underlying assumptions, the SAR models and the GWR.

First, the SAR models correct for the spatial autocorrelation in the residuals of the regression by including an autoregressive parameter in the model. In particular, three main variations of the SAR models exist depending on where the autoregressive process is applied on; on the error term (spatial error model), the dependent variable (spatial lag model), and on the independent variables (spatial lag of X). In order to determine which model to estimate for the particular case at hand, all models are estimated and are evaluated on their ability to resolve the spatial autocorrelation issues. Spatial autocorrelation is measured by the local Moran's I index, which is a measure of correlation, which takes into account the spatial structure of data imposed by a spatial weight matrix (neighborhood matrix). Initially, the necessity to proceed to the estimation of SAR models is assessed by calculating the spatial autocorrelation of the OLS residuals, which if present, leads to biased estimated coefficients. Spatial autocorrelation is present and significant, having values close to 0.1 for different tested definitions of neighborhood. Driven by this, the necessity to proceed to the estimation of SAR models arises, in order to resolve that issue and obtain consistent and unbiased coefficient estimations. In order to identify which spatial models resolves the spatial autocorrelation issues, we make use of the Lagrange Multiplier tests to check for error dependence, or/ and missing lagged dependent variable (Anselin et. al, 1996). The results of the tests are presented in and, where they exhibit the existence of error dependence and none missing lagged variable, for both spatial analysis levels. Based on these findings, we proceed to the estimation of the spatial error models (SER), which are found to be able to fully resolve spatial autocorrelation issues.

Table 10: Lagrange Multiplier tests for the nationwide dataset

	Year 2000		Year 2005		Year 2010	
Lagrange Multiplier Tests	Estimate	Pr(>|t|)	Estimate	Pr(>|t|)	Estimate	Pr(>|t|)
LMError	24.78	***	46.39	***	56.78	***
Lmlag	10.22	*	2.59	.	0.37	
RLMerror	26.15	***	44.53	***	56.42	***
RLMlag	11.59	***	0.74		0.08	
*Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1*						

Table 11: Lagrange Multiplier tests for the agglomeration areas

	Year 2000		Year 2005		Year 2010	
Lagrange Multiplier Tests	Estimate	Pr(>|t|)	Estimate	Pr(>|t|)	Estimate	Pr(>|t|)
LMError	21.61	***	9.16	***	4.43	*
LMlag	0.92		0.15		0.55	
RLMerror	21.19	***	9.42	**	4.19	*
RLMlag	0.50		0.42		0.31	
*Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1*						

Spatial error models, which assume that the spatial dependence is in the error term of the model, are found to be the one able to resolve that issue and thus they are the ones reported. The formulation of such models is:

$$Y = \beta X + u \ (2)$$

$$with \ \ u = \lambda W u + \varepsilon$$

where Y is a vector with N values of the dependent variable, β is a vector with the regression coefficients, X is a matrix with the independent variables, u the error term, λ the spatial autoregressive coefficient, W a matrix with the spatial structure having dimensions $N \times N$, and ε a vector of independent and identically distributed (iid) error terms.

A key aspect of the spatial regression models is to determine the spatial structure of the data. This is facilitated by the inclusion of a spatial weight matrix in the model formulation. Thereupon, the spatial weight matrix incorporates in the model information about the extent of the neighborhood, the type of the adjacency, and the relative weight that should be assigned on the neighboring locations.

The employed spatial matrix is determined through an iterative process of identifying the existence of spatial autocorrelation in the OLS residuals, and then on its ability to account for it properly in the SAR models and on the basis of minimizing the Akaike criterion value (goodness-of-fit measure). More specifically, the second order spatial contiguity matrix is found to be the optimum one for the case at hand, while spatial weight matrices based on the Euclidean distance and also the generalized cost (as a measure of network distance) are tested as well but yield worse results. The same correction for the error term is applied also to the panel data models.

The results of the different models mentioned above are presented in the following sections. Spatial error models and GWR are estimated making use of the spdep package in R (Bivand et al., 2011).

4.3.1 Spatial error models

Table 12: Spatial error models

	Year 2000		Year 2005		Year 2010	
Independent Variable: Ln mean salary	Estimate	Pr(>|t|)	Estimate	Pr(>|t|)	Estimate	Pr(>|t|)
Intercept	6.487	***	6.916	***	6.900	***
Ln car accessibility	0.010	**	0.017	***	0.011	**
Ln public transport accessibility	0.016	***	0.013	***	0.012	**
Ln number of local employed	0.017	***	0.011	***	0.014	***
Commuters from outside Switzerland	-0.117	***	-0.095	***	-0.097	***
Short residence permit	-0.198	**	-0.136	***	-0.210	**
Average duration in-post	0.003	**	0.007	***	0.005	***
Ln average age	0.345	***	0.278	***	0.318	***
Men	0.180	***	0.063	***	0.134	***
Tertiary education	0.891	***	0.685	***	0.582	***
Professional training	0.533	***	0.227	***	0.333	***
Further vocational training	0.216	***	0.189	***	0.247	***
Teaching degree	0.181	*	0.196	***	0.324	***
Highschool diploma	0.644	***	0.240	**	0.236	**
Vocational training	0.071	***	0.034	*	0.020	
Positions with highest demands	0.429	***	0.375	***	0.364	***
Positions with qualified indep. work	0.195	***	0.246	***	0.229	***
Positions with professional skills	0.134	***	0.192	***	0.136	***
Working (3rd sector)	0.201	***	0.160	***	0.067	*
Working (private sector)	-0.110	***	-0.106	***	-0.067	***
Working (manufacturing)	-0.229	***	-0.249	***	-0.102	***
Working (FIRE)	0.139	***	-0.005		0.074	.
Working (hotel, restaurants)	-0.133	***	-0.131	***	-0.112	***
lamda parameter	0.218	***	0.282	***	0.303	***
AIC		-2696		-4688		-4187
AIC ols		-2676		-4651		-4143
Nagelkerke pseudo-R-squared		0.686		0.656		0.615
Residuals' spatial autocorrelation	-0.002		-0.005		-0.003	
OLS residuals' spatial autocorrelation	0.07	***	0.057	***	0.065	***
# observations		1448		2298		2229

*Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1*

Table 13: SER models for the agglomeration areas

	Year 2000		Year 2005		Year 2010	
Independent Variable: Ln mean salary	Estimate	Pr(>|t|)	Estimate	Pr(>|t|)	Estimate	Pr(>|t|)
Intercept	6.311	***	6.261	***	5.856	***
Ln car accessibility	0.028	***	0.021	***	0.027	***
Ln public transport accessibility	0.017	**	0.037	***	0.026	***
Ln number of local employed	0.019	***	0.010	***	0.016	***
Commuters from outside Switzerland	-0.132	***	-0.081	**	-0.104	***
Short residence permit	-0.093		0.184	*	0.521	***
Average duration in-post	0.004	**	0.007	***	0.003	.
Ln average age	0.326	***	0.322	***	0.452	***
Men	0.141	***	-0.007		0.069	*
Tertiary education	0.876	***	0.727	***	0.593	***
Professional training	0.451	***	0.349	***	0.248	***
Further vocational training	0.249	***	0.203	***	0.251	***
Teaching degree	0.422	***	0.298	***	0.469	***
Highschool diploma	0.645	***	0.463	***	0.374	**
Vocational training	0.074	**	0.107	***	0.055	.
Positions with highest demands	0.278	**	0.464	***	0.652	***
Positions with qualified indep. work	0.310	***	0.210	***	0.289	***
Positions with professional skills	0.130	***	0.112	***	0.194	***
Working (3rd sector)	0.221	**	0.376	***	0.261	***
Working (private sector)	-0.063	**	-0.070	***	-0.044	*
Working (manufacturing)	-0.306	***	-0.185	***	-0.019	
Working (FIRE)	0.144	***	0.159	*	0.204	**
Working (hotel, restaurants)	-0.158	***	-0.109	***	-0.134	***
lamda parameter	0.200	***	0.128	**	0.105	**
AIC		-1596		-2008		-1865
AIC ols		-1576		-2001		-1862
Nagelkerke pseudo-R-squared		0.748		0.694		0.679
Residuals' spatial autocorrelation	-0.010		-0.005		-0.006	
OLS residuals' spatial autocorrelation	0.138	***	0.078	***	0.054	*
# observations		789		958		954

*Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1*

4.3.2 Spatial panel data models

In line with the previous models, the next category of models is the spatial error panel data models, according to the formulation and the estimation techniques presented by Millo and Piras (2012). The existence of spatial dependence in the error terms is tested by calculating a conditional Lagrange multiplier test which shows statistically significant spatial dependence (see Baltagi et. al, 2003). The corresponding results are presented in the following tables, for the two levels of spatial analysis and the two time periods (Table 14 – 17).

Table 14: Spatial panel data models

Independent variable: Ln mean salary	SER pooled		SER with TE	
	Estimate	Pr(>|t|)	Estimate	Pr(>|t|)
Intercept	5.388	***	6.257	***
Year 2005 dummy (time-effect)			0.081	***
Year 2010 dummy (time-effect)			0.118	***
Ln car accessibility	0.015	***	0.012	***
Ln public transport accessibility	0.009	**	0.017	***
Ln number of local employed	0.018	***	0.015	***
Commuters from outside Switzerland	-0.055	***	-0.097	***
Short residence permit	-0.056		-0.146	***
Average duration in-post	0.000		0.003	***
Ln average age	0.672	***	0.406	***
Men	0.136	***	0.140	***
Tertiary education	0.766	***	0.759	***
Professional training	0.408	***	0.371	***
Further vocational training	0.310	***	0.232	***
Teaching degree	0.362	***	0.346	***
Highschool diploma	0.420	***	0.341	***
Vocational training	0.073	***	0.070	***
Positions with highest demands	0.636	***	0.448	***
Positions with qualified indep. work	0.378	***	0.244	***
Positions with professional skills	0.239	***	0.166	***
Working (3rd sector)	0.065	*	0.182	***
Working (private sector)	-0.069	***	-0.077	***
Working (manufacturing)	-0.176	***	-0.211	***
Working (FIRE)	0.018		0.134	***
Working (hotel, restaurants)	-0.129	***	-0.120	***
Rho	0.241	***	0.277	***
Balanced panel observations			1374 (total = 4122)	

*Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1*

Table 15: Spatial panel data models for the agglomeration areas

Independent variable: Ln mean salary	SER pooled		SER with TE	
	Estimate	Pr(>|t|)	Estimate	Pr(>|t|)
Intercept	5.363	***	6.182	***
Year 2005 dummy (time-effect)			0.082	***
Year 2010 dummy (time-effect)			0.109	***
Ln car accessibility	0.030	***	0.025	***
Ln public transport accessibility	0.017	***	0.024	***
Ln number of local employed	0.017	***	0.014	***
Commuters from outside Switzerland	-0.080	***	-0.119	***
Short residence permit	0.125		0.058	
Average duration in-post	0.001		0.004	***
Ln average age	0.598	***	0.343	***
Men	0.100	***	0.085	***
Tertiary education	0.699	***	0.695	***
Professional training	0.351	***	0.333	***
Further vocational training	0.271	***	0.187	***
Teaching degree	0.443	***	0.428	***
Highschool diploma	0.586	***	0.432	***
Vocational training	0.082	***	0.086	***
Positions with highest demands	0.702	***	0.552	***
Positions with qualified indep. work	0.449	***	0.303	***
Positions with professional skills	0.242	***	0.163	***
Working (3rd sector)	0.101	*	0.259	***
Working (private sector)	-0.031	*	-0.034	**
Working (manufacturing)	-0.138	**	-0.210	***
Working (FIRE)	0.049		0.172	***
Working (hotel, restaurants)	-0.166	***	-0.159	***
Rho	0.224	***	0.277	***
Balanced panel observations			1374 (total = 4122)	

*Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

Table 16: Spatial panel data models for years 2005-2010

Independent variable: Ln mean salary	SER pooled		SER with TE	
	Estimate	Pr(>|t|)	Estimate	Pr(>|t|)
Intercept	6.674	***	6.863	***
Year 2010 dummy (time-effect)			0.039	***
Ln car accessibility	0.025	***	0.016	***
Ln public transport accessibility	0.002		0.012	***
Ln number of local employed	0.013	***	0.012	***
Commuters from outside Switzerland	-0.081	***	-0.087	***
Short residence permit	-0.201	***	-0.165	***
Average duration in-post	0.006	***	0.006	***
Ln average age	0.354	***	0.301	***
Men	0.098	***	0.107	***
Tertiary education	0.647	***	0.623	***
Professional training	0.318	***	0.302	***
Further vocational training	0.246	***	0.238	***
Teaching degree	0.331	***	0.304	***
Highschool diploma	0.347	***	0.335	***
Vocational training	0.050	***	0.039	**
Positions with highest demands	0.394	***	0.399	***
Positions with qualified indep. work	0.268	***	0.235	***
Positions with professional skills	0.181	***	0.165	***
Working (3rd sector)	0.094	***	0.086	***
Working (private sector)	-0.073	***	-0.081	***
Working (manufacturing)	-0.175	***	-0.169	***
Working (FIRE)	0.016		0.038	
Working (hotel, restaurants)	-0.089	***	-0.105	***
Rho	0.293	***	0.303	***
Balanced panel observations			2117 (total 4234)	

*Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1*

Table 17: Spatial panel data models for years 2005-2010 for the agglomeration areas

Independent variable: Ln mean salary	SER pooled		SER with TE	
	Estimate	Pr(>|t|)	Estimate	Pr(>|t|)
Intercept	5.858	***	6.060	***
Year 2010 dummy (time-effect)			0.033	***
Ln car accessibility	0.036	***	0.025	***
Ln public transport accessibility	0.022	***	0.029	***
Ln number of local employed	0.013	***	0.013	***
Commuters from outside Switzerland	-0.084	***	-0.088	***
Short residence permit	0.164	*	0.251	**
Average duration in-post	0.004	***	0.005	***
Ln average age	0.444	***	0.395	***
Men	0.043	.	0.048	*
Tertiary education	0.626	***	0.613	***
Professional training	0.296	***	0.298	***
Further vocational training	0.236	***	0.234	***
Teaching degree	0.448	***	0.435	***
Highschool diploma	0.347	***	0.360	***
Vocational training	0.067	**	0.073	***
Positions with highest demands	0.542	***	0.542	***
Positions with qualified indep. work	0.274	***	0.235	***
Positions with professional skills	0.164	***	0.141	***
Working (3rd sector)	0.274	***	0.278	***
Working (private sector)	-0.040	**	-0.048	***
Working (manufacturing)	-0.148	***	-0.134	***
Working (FIRE)	0.147	**	0.174	***
Working (hotel, restaurants)	-0.095	***	-0.110	***
Rho	0.303	***	0.319	***
Balanced panel observations			2117 (total 4234)	

*Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1*

4.4 GWR

The next category of spatial models is the GWR approach which treats the issue of the spatial heterogeneity by providing localized estimates of the coefficients and investigates how they vary over space, making the assumption that the structural equation is not holding over space for various reasons. In particular, GWR explores spatial non-stationarity for given bandwidth value, either global (in the form of fixed distance) or adaptive (in the form of k-nearest neighbors), and a gauss weighting distance function. The bandwidth value defines the extent of the neighborhood and it is calculated on the basis of minimizing a cross-validation function. In the present study, the choice of an adaptive bandwidth is qualified in order to allow accounting for the different spatial density of zones. In this particular case, the adaptive bandwidth, in the form of k-nearest neighbors (operationalized as the portion out of the total observations), is found to take values of 10.9%, 7.52%, and 7.58% respectively for each year. The decrease in the bandwidth accrues from the increased size of the datasets in the subsequent years following 2000 while the actual optimum number of k-nearest neighbors remains almost constant across the years. The reported results per year correspond to a range of values along with their distribution, while treating for the spatial heterogeneity also resolves partially spatial autocorrelation issues (Table 18 - 19). In addition, a visual representation of the spatial variation of the public transport accessibility coefficient is given in the accompanying plot (Figure 5). It should be noted than an approved way of estimating models assuming combined spatial heterogeneity and serial autocorrelation (panel data) is still not present in the econometric literature and thus is not tested.

Table 18: GWR results, Year 2000

Independent Variable: Ln mean salary	Min.	1st Quart.	Median	3rd Quart.	Max.	Global
Intercept	5.460	6.098	6.405	6.552	7.198	6.495
Ln car accessibility	0.004	0.011	0.016	0.021	0.029	0.010
Ln public transport accessibility	-0.004	0.009	0.018	0.027	0.032	0.019
Ln number of local employed	0.004	0.012	0.018	0.021	0.028	0.017
Commuter from outside Switzerland	-0.183	-0.083	-0.032	0.125	0.254	-0.112
Short residence permit	-0.772	-0.342	-0.195	-0.103	0.145	-0.185
Average duration in-post	-0.004	0.001	0.002	0.004	0.006	0.003
Ln average age	0.248	0.314	0.351	0.420	0.548	0.338
Men	0.078	0.188	0.211	0.237	0.267	0.178
Tertiary education	-0.020	0.646	0.754	0.972	1.306	0.898
Professional training	0.273	0.406	0.462	0.626	0.831	0.523
Further vocational training	0.007	0.160	0.214	0.261	0.373	0.210
Teaching degree	-0.239	0.152	0.320	0.411	0.731	0.167
Highschool diploma	-0.187	0.390	0.601	0.833	1.124	0.624
Vocational training	0.001	0.037	0.049	0.067	0.080	0.063
Positions with highest demands	0.082	0.322	0.485	0.665	1.085	0.434
Positions with qualified indep. work	0.060	0.110	0.201	0.255	0.321	0.206
Positions with professional skills	0.064	0.100	0.134	0.158	0.205	0.144
Working (3rd sector)	-0.169	0.102	0.190	0.333	0.432	0.199
Working (other private sector)	-0.167	-0.138	-0.117	-0.088	-0.052	-0.116
Working (manufacturing)	-0.375	-0.278	-0.239	-0.180	0.024	-0.224
Working (FIRE)	0.087	0.113	0.141	0.190	0.279	0.146
Working (hotel, restaurants)	-0.161	-0.127	-0.090	-0.065	0.027	-0.139
Local adjusted R-squared	0.702	0.724	0.736	0.745	0.771	0.787
Residuals' spatial autocorrelation					0.038	**
OLS residuals' spatial autocorrelation					0.113	***

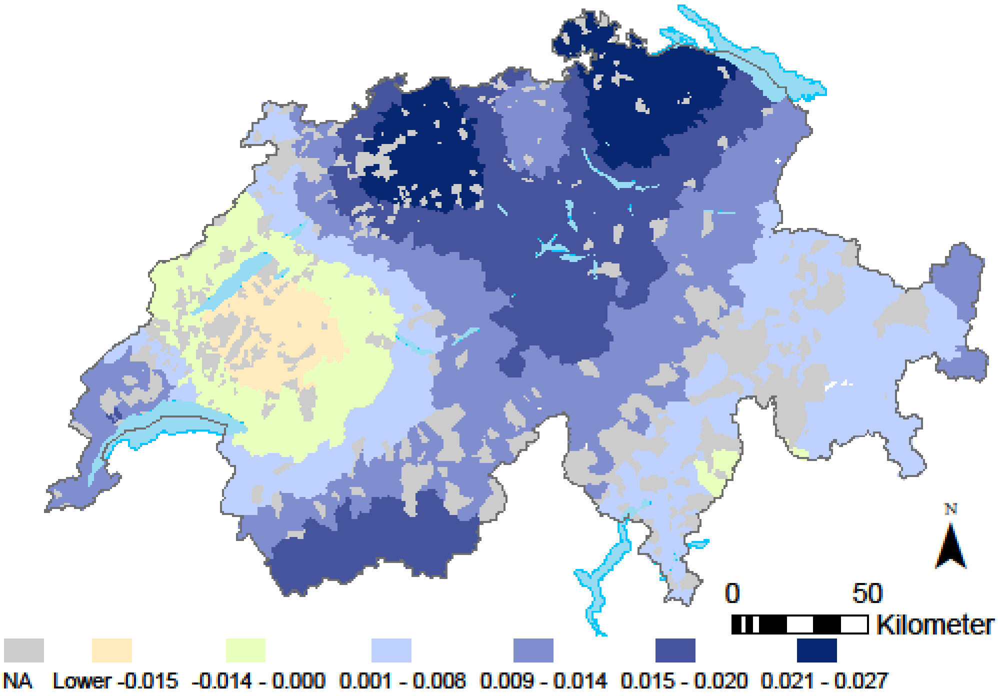
Table 19: GWR results, Year 2005

Independent Variable: Ln mean salary	Min.	1st Quart.	Median	3rd Quart.	Max.	Global
Intercept	6.313	6.509	6.768	7.030	7.383	6.923
Ln car accessibility	0.018	0.020	0.021	0.021	0.022	0.017
Ln public transport accessibility	0.004	0.015	0.021	0.027	0.030	0.016
Ln number of local employed	0.007	0.008	0.009	0.009	0.011	0.012
Commuter from outside Switzerland	-0.016	0.069	0.114	0.146	0.160	-0.087
Short residence permit	-0.150	-0.135	-0.120	-0.100	-0.057	-0.147
Average duration in-post	0.006	0.007	0.007	0.008	0.009	0.007
Ln average age	0.114	0.208	0.262	0.307	0.347	0.267
Men	0.017	0.049	0.067	0.087	0.116	0.059
Tertiary education	0.546	0.570	0.582	0.605	0.654	0.691
Professional training	0.219	0.285	0.327	0.351	0.398	0.220
Further vocational training	0.275	0.292	0.313	0.329	0.357	0.187
Teaching degree	0.083	0.201	0.244	0.276	0.318	0.192
Highschool diploma	0.205	0.276	0.316	0.354	0.404	0.236
Vocational training	-0.042	-0.038	-0.033	-0.027	-0.016	0.035
Positions with highest demands	0.383	0.456	0.499	0.567	0.659	0.408
Positions with qualified indep. work	0.081	0.132	0.152	0.176	0.202	0.255
Positions with professional skills	0.149	0.157	0.166	0.181	0.205	0.200
Working (3rd sector)	0.101	0.140	0.173	0.219	0.268	0.157
Working (other private sector)	0.001	0.010	0.014	0.016	0.018	-0.109
Working (manufacturing)	-0.345	-0.327	-0.320	-0.308	-0.291	-0.245
Working (FIRE)	0.063	0.124	0.176	0.245	0.291	-0.010
Working (hotel, restaurants)	-0.161	-0.134	-0.120	-0.110	-0.101	-0.127
Local adjusted R-squared	0.609	0.707	0.730	0.728	0.745	0.807
Residuals' spatial autocorrelation					0.027	***
OLS residuals' spatial autocorrelation					0.103	***

Table 20: GWR results, Year 2010

Independent Variable: Ln mean salary	Min.	1st Quart.	Median	3rd Quart.	Max.	Global
Intercept	5.722	6.716	6.957	7.163	7.703	6.864
Ln population accessibility	0.000	0.010	0.016	0.027	0.039	0.011
Ln public transport accessibility	-0.030	0.005	0.012	0.018	0.027	0.015
Ln number of local employed	0.006	0.011	0.013	0.015	0.019	0.014
Commuter from outside Switzerland	-0.248	-0.109	0.005	0.060	0.207	-0.097
Short residence permit	-1.069	-0.645	-0.507	-0.288	0.112	-0.189
Average duration in-post	-0.001	0.003	0.004	0.006	0.010	0.005
Ln average age	0.111	0.247	0.304	0.378	0.523	0.319
Men	0.077	0.117	0.139	0.160	0.217	0.126
Tertiary education	0.377	0.494	0.535	0.613	0.733	0.594
Professional training	0.098	0.201	0.279	0.364	0.468	0.317
Further vocational training	-0.010	0.159	0.215	0.273	0.483	0.233
Teaching degree	0.104	0.233	0.297	0.405	0.516	0.321
Highschool diploma	-0.582	-0.129	0.026	0.202	0.723	0.253
Vocational training	-0.172	-0.099	-0.034	0.034	0.146	0.020
Positions with highest demands	0.143	0.310	0.465	0.602	0.781	0.397
Positions with qualified indep. work	0.067	0.204	0.247	0.343	0.481	0.242
Positions with professional skills	0.013	0.115	0.150	0.211	0.385	0.145
Working (3rd sector)	-0.145	0.030	0.068	0.093	0.170	0.071
Working (other private sector)	-0.145	-0.116	-0.100	-0.051	0.086	-0.071
Working (manufacturing)	-0.327	-0.223	-0.176	-0.108	-0.062	-0.102
Working (FIRE)	-0.100	0.071	0.131	0.171	0.255	0.061
Working (hotel, restaurants)	-0.239	-0.147	-0.078	-0.028	0.102	-0.108
Local adjusted R-squared	0.567	0.654	0.676	0.680	0.708	0.743
Residuals' spatial autocorrelation					0.040	***
OLS residuals' spatial autocorrelation					0.097	***

Figure 5: GWR Estimated coefficient of public transport accessibility in year 2010



5. Conclusion

In the present study a quantification of the agglomeration effects that arise by the improvement of the public transport supply is attempted. Different modelling approaches are tested to check the stability of the results over the modelling approaches. Improvements on public transport accessibility are found to be able having a positive impact on the productivity of Switzerland. The corresponding elasticity values are found to lie in the range of 1 to 3% for the nationwide spatial analysis level, while for the agglomeration areas the effects are significantly higher and take values between 2 and 4%. Moreover, elasticity values exhibit variation over the analysis period reflecting a differentiated dynamic of the externalities that can result due to the public transport improvements. The results are summarized in Table 21 and Table 22.

Table 21: Summary of estimated elasticity values for public transport accessibility

	Estimated Public transport elasticity		
Model	Year 2000	Year 2005	Year 2010
OLS	1.80%	1.60%	1.50%
Spatial error	1.60%	1.30%	1.20%
Pooled OLS		1.20%	
Pooled OLS for 2005-2010		0.7% (insignificant)	
Time-effects		2.00%	
Time-effects for 2005-2010		1.50%	
SER pooled OLS		0.90%	
SER pooled OLS for 2005-2010		0.2% (insignificant)	
SER with time-effects		1.70%	
SER with time-effects for 2005-2010		1.20%	
GWR	-0.4 - 3.25%	0.4 - 3%	-0.3 - 2.7%

Table 22: Summary of estimated elasticity values for public transport accessibility for the urban agglomeration areas

	Estimated Public transport elasticity		
Model	Year 2000	Year 2005	Year 2010
OLS	2.10%	3.90%	2.70%
Spatial error	1.70%	3.70%	2.60%
Pooled OLS		1.80%	
Pooled OLS for 2005-2010		2.60%	
Time-effects		2.70%	
Time-effects for 2005-2010		3.30%	
SER pooled OLS		1.70%	
SER pooled OLS for 2005-2010		2.20%	
SER with time-effects		2.40%	
SER with time-effects for 2005-2010		2.90%	

In addition, GWR is used to investigate the spatial heterogeneity of the assumed causal relationships. Interestingly, as it can be seen in Figure 5, the higher values of the accessibility parameter are found to occur in the north-eastern part of Switzerland, and in the south-western part (Geneva region). In the case of the agglomeration effect due to the employment density, operationalized by the local employment positions per zone, the corresponding elasticity values are found to lie in the range of 1to 2%.

Acknowledgment

This research is in collaboration with ecoplan AG and funded by the SBB Research Lab.

6. References

Anselin, L. (1988) Spatial Econometrics: Methods and Models, *Studies in Operational Regional Science*, Springer Science+Business Media, Dordrecht.

Anselin, L., A. K. Bera, R. Florax,. and M. J. Yoon (1996) Simple diagnostic tests for spatial dependence, *Regional Science and Urban Economics*, **26** 77–104.

Baltagi, B.H., S.H. Song, and W. Koh (2003) Testing panel data regression models with spatial error correlation. *Journal of Econometrics*, **117** 123–150.

Bivand, R., L. Anselin, O. Berke, A. Bernat, M. Carvalho, Y. Chun, C. F. Dormann et al. (2011) spdep: Spatial dependence: weighting schemes, statistics and models.

Breusch, T.S. and A.R. Pagan (1980) The Lagrange multiplier test and its applications to model specification in econometrics, *Review of Economic Studies*, **47** 239–253.

Breusch, T.S. and A.R. Pagan (1979) A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica*, **47** 1287–1294.

Charlton, M. and S. Fotheringham (2009) Geographically Weighted Regression, White paper, National Centre for Geocomputation. *National University of Ireland Maynooth*.

Chatman, D. G. and R. B. Noland (2011) Do Public Transport Improvements Increase Agglomeration Economies? A Review of Literature and an Agenda for Research, *Transport Reviews*, **31**(6), 725–742.

Charlton, M. and S. Fotheringham (2009) Geographically Weighted Regression, White paper, *National Centre for Geocomputation. National University of Ireland Maynooth*,.

Chatman, D. G. and R. B. Noland (2011) Do Public Transport Improvements Increase Agglomeration Economies? A Review of Literature and an Agenda for Research, *Transport Reviews*, 31 (6) 725–742.

Graham, D. (2007) Agglomeration, productivity and transport investment, *Journal of transport economics and policy (JTEP)*, 41 (August 2006) 317–343.

Graham, D. and K. Van Dender (2011) Estimating the agglomeration benefits of transport investments: some tests for stability, *Transportation*, 2010 (2009).

Halás, M. P. Klapka and P. Kladivo (2014) Distance-decay functions for daily travel-to-work flows, *Journal of Transport Geography*, 35 107–119.

LeSage, J. and R. Pace (2004) Spatial and spatiotemporal econometrics, *Advances in Econometrics*, 18 (04) 1–32.

Millo, G. and G. Piras (2012) splm: Spatial Panel Data Models in R, *Journal of Statistical Software*, VV (Ii).

Halás, M. P. Klapka and P. Kladivo (2014) Distance-decay functions for daily travel-to-work flows, *Journal of Transport Geography*, **35** 107–119.

Hansen, W. G. (1959) How Accessibility Shapes Land Use, Journal of the American Institute of Planners, **25**(2) 73-76.

Hausman, J.A. (1978), Specification tests in econometrics, *Econometrica*, **46** 1251–1271.

Honda, Y. (1985) Testing the error components model with non–normal disturbances, *Review of Economic Studies*, **52** 681–690.

Koenker R. (1981) A Note on Studentizing a Test for Heteroscedasticity. *Journal of Econometrics* **17** 107–112.

Kwan, M.-P. (1998) Space-Time and Integral Measures of Individual Accessibility: A Comparative Analysis Using a Point-based Framework, *Geographical Analysis*, **30**(3) 191-216.

LeSage, J. and R. Pace (2004) Spatial and spatiotemporal econometrics, *Advances in Econometrics*, 18 (04) 1–32.

Millo, G. and G. Piras (2012) splm: Spatial Panel Data Models in R, *Journal of Statistical Software*, VV (Ii).

R Development Core Team (2011) R: A Language and Environment for Statistical Computing, *R Foundation for Statistical Computing*, Vienna, Austria.

White H. (1980) A Heteroskedasticity-Consistent Covariance Matrix and a Direct Test for Heteroskedasticity, *Econometrica* **48** 817–838.