Robustness of statistical models for the relationship between accessibility and regional development – A case study of Swedish municipalities

Ida Kristoffersson*[1], Jonas Westin[2]

[1] Docent, Department of Communication, Quality Management, and Information Systems, Mid Sweden University, Sweden, ida.kristoffersson@miun.se

[2] Associate professor, Department of Mathematics and Mathematical Statistics, Umeå University, Sweden, jonas.westin@umu.se

SHORT SUMMARY

This paper examines the relationship between car accessibility and regional growth across all 290 Swedish municipalities, using accessibility data from the PIPOS platform for 2014 and 2020 and population and gross regional product data for 2020–2023. We estimate a range of regression models, varying the inclusion of control variables, the treatment of outliers, and the separate estimation for municipalities in high-growth and low-growth counties. Simple linear models show a strong positive correlation between accessibility and regional growth. However, this relationship is not robust: adding control variables reverses the effect of accessibility, removing outliers severely weakens the explanatory power, and the relationship disappears entirely in municipalities in low-growth regions. Our main conclusion is that researchers using statistical approaches to analyse the relationship between accessibility and regional development should demonstrate that their results are robust across model specifications, as conclusions can depend heavily on methodological choices.

Keywords: Accessibility, transport investment, regional development, population change

1. INTRODUCTION

Urbanization is a global phenomenon: from 1950 to 2025, the global share of people living in cities of at least 50,000 inhabitants has increased from 20 to 45 percent (United Nations 2025). Sweden is no exception to this trend. Although the country's population has increased by 2.3 million from 1980 to 2024, growth has been concentrated in larger urban areas, while 47% of Sweden's 290 municipalities have experienced population decline (SCB 2026). In general, it is the already large municipalities (in terms of population, not geographical size) that have grown, and the small ones that have declined even further. The size and composition of the population are the most important planning prerequisites for Swedish municipalities (Syssner and Olausson 2016). United Nations (2025) states three main forces that drive urbanization: economic opportunities, demographic shifts and infrastructure development. However, it is still not clear what role infrastructure development has in this pattern of urbanization over time.

The long-term population and regional economic effects of a transport investment are traditionally modelled using regression analysis based on an approach developed by Carlino and Mills (1987), in which the long-term location of households depends on factors in the regional environment, such as access to workplaces. Time has, however, shown a large variation in the results and implications of Carlino and Mills' model types, which are sensitive to geographical location, spatial resolution, and model specification, and these types of models have been criticized, for example,

in Rickman (2010). Using Carlino and Mills' types of models, researchers have failed to find a clear answer to whether jobs follow people or people follow jobs (Hoogstra et al. 2017).

The Carlino and Mills approach is, however, still widely used. For example it is the basis for the Swedish SAMLOK model (Inregia AB 2001, 2003) developed in the 90s and still used by The Swedish Transport Administration to estimate location effects of transport investments, even though it has in recent years been re-estimated based on more recent data (1995 to 2006 instead of 1985 to 1997) (Anderstig et al. 2021; Berglund 2025), and it is also often utilized in grey literature for calculation of regional effects of air accessibility (Sweco 2025; WSP 2010; WSP and ÅF 2014).

The approach assumes uniform effects across regions, overlooking that the same improvement can reinforce agglomeration in large regions while causing a decline in smaller ones. This contrasts with New Economic Geography (Fujita et al. 1999), which emphasises that regional differences can lead to divergent regional outcomes. Furthermore, the direction of causality between accessibility and population changes is not clear from previous studies: are transport investments made in areas where housing developments are already planned, or are housing developments planned because of population increases due to increased accessibility by transport investments? This endogeneity problem, essentially a chicken-and-egg issue, is difficult to untangle.

In this paper, we demonstrate that the relationship between transport accessibility and regional outcomes in terms of population and economic development is more complex than what simple regression models can capture, and that conclusions drawn from models assuming constant parameters regardless of local conditions are not robust. We thus show that results from such statistical models, particularly those that assume increased accessibility inevitably leads to regional growth, are not robust and should therefore not be used as a basis for policy recommendations. This implies, for example, that a transport investment cannot be expected to generate the same population or economic effects in high-growth and low-growth regions.

2. METHODOLOGY

We use accessibility data from a platform called PIPOS (https://pipos.se/) developed by a cooperation between The Swedish Agency for Economic and Regional Growth and The Swedish Transport Administration. In PIPOS, accessibility is calculated for each 250mx250m grid where there is population. Accessibility is calculated for 43 different types of so-called "travel reasons", for example, accessibility to workplaces in the construction business. For each travel reason, accessibility is calculated using an index, which considers both how easy it is to reach the destination zone and how large the supply is for this travel reason at the destination zone, for all destination zones within a certain maximum travel time from the origin. This index is calculated according to Equation (1) and then z-transformed. The accessibility index A_{ir} for origin zone i and travel reason r is calculated as:

$$A_{ir} = \sum_{j \epsilon J} R_j \, 9^{\frac{T}{t_{ij}}} \qquad (1)$$

where R_j is the number of occurrences of travel reason r in destination zone j, J is the set of destination zones which lie within the maximum travel time T, and t_{ij} is the fastest travel time between origin zone i and destination zone j. The accessibility index is calculated for the years 2014 and 2020, for car as travel mode. Weighing with population, we aggregate this accessibility index over origin zones i to municipality level. Accessibility is aggregated across all travel reasons r into A, with variables expressed as z-scores, i.e., A is the total accessibility in 2020, whereas

ΔA is the difference in accessibility between year 2014 and 2020. From Statistics Sweden, we have data on population, gross regional product, and socioeconomic variables for each municipality, which are used as dependent and control variables in the estimations, see Table 1.

We estimate and compare a number of regression models for how regional development in the form of changes in population or change in gross regional product between 2020 and 2023 (ΔB / ΔG) could be related to accessibility, changes in accessibility, and socioeconomic control variables. Equation (2) shows the simplest model specification in which accessibility levels in 2020 are assumed to be the only explanatory variable for population change, included together with an intercept and an error term. Equation (3) is a similar relationship, but with a change in accessibility between 2014 and 2020 as explanatory variable, and all available control variables added.

$$\Delta B = \beta_0 + \beta_1 A + \varepsilon \tag{2}$$

$$\Delta B = \beta_0 + \beta_2 \Delta A + \beta_3 D + \beta_4 I + \beta_5 G_i + \beta_6 G_e + \beta_7 G + \beta_8 G_{ih} + \beta_9 G_{eh} + \beta_{10} G_h + \beta_{11} B_h + \beta_{12} B + \beta_{13} E + \beta_{14} H + \varepsilon \tag{3}$$

Table 1: Variable notation in the estimations.

Variable	Notation	Year
Dependent variables		
Change in population	ΔB	2020-2023
Change in gross regional product	ΔG	2020-2023
Independent variables		
Accessibility	A	2020
Change in accessibility	ΔA	2014-2020
Average age	D	2020
Average income	I	2020
Gross regional product per inhabitant	G_i	2020
Gross regional product per employed	G_e	2020
Gross regional product	G	2020
Historic change in gross regional product per inhabitant	G_{ih}	2014-2020
Historic change in gross regional product per employed	G_{eh}	2014-2020
Historic change in gross regional product	G_h	2014-2020
Historic population change	B_h	2014-2020
Population	B	2020
Share of employed	E	2020
Share of highly educated	H	2020

Several models are estimated and compared. The models are compared along five different aspects: 1) not including control variables, including all control variables, or adding control variables stepwise, 2) removing or not removing outliers, 3) using accessibility or change in accessibility as explanatory variable, 4) using change in population or change in gross regional product as dependent variable, and 5) separate estimation for municipalities in high-growth and low-growth counties. Outliers are identified and removed using the interquartile range (IQR) method. Forward stepwise selection with 5-fold cross-validated R^2 is used to select controls, guarding against overfitting while keeping the accessibility variable fixed to allow direct comparability across models. Municipalities are also grouped into high-growth (top 11) and low-growth (bottom 10) counties, based on aggregated change in the respective dependent variable. Model fit is evaluated using R^2, adjusted R^2, AIC, BIC, and F-statistic, and parameter significance levels.

3. RESULTS AND DISCUSSION

Results of model estimations for population change and gross regional product as dependent variables are summarized in Tables 2 and 3, respectively. Simple linear models with only accessibility or a change in accessibility as explanatory variable show a strong positive correlation with both population and gross regional product growth, both with and without outliers. However, this positive relationship does not hold up under closer examination.

Regarding aspect 1, adding control variables, either all at once or stepwise, has a profound effect: with outliers present, the accessibility parameters remain significant but turn negative, whereas with outliers removed, they become entirely insignificant. This suggests that the effect of accessibility in simple models is confounded by underlying regional characteristics; municipalities with high accessibility also tend to have favourable historic growth trends and conditions that drive regional development.

Regarding aspect 2, removing outliers substantially reduces the models' explanatory power. For population change, R^2 drops from around 30% to 10% without controls, and from around 90% to 60% with controls. For gross regional product, the pattern is similar but more pronounced, with R^2 dropping from around 60% to 10% without controls, and from around 96% to 60% with controls, highlighting that the outliers have a large impact on the explanatory power of the models.

These patterns are similar whether accessibility levels or changes in accessibility are used as explanatory variables (aspect 3), and regardless of whether population or gross regional product is used as the dependent variable (aspect 4). However, the gross regional product results can be seen as an amplified version of the population results.

Regarding aspect 5, Figure 1 shows the results of estimating models separately for municipalities in high-growth and low-growth counties. For municipalities in high-growth counties, the results are very similar to those for all municipalities. However, for municipalities in low-growth counties, even the simple, previously strong relationship between accessibility and regional growth turns insignificant. This is consistent with the New Economic Geography argument: accessibility improvements appear to reinforce agglomeration in already growing regions rather than generating growth in declining ones.

Table 2: Results of model estimations of population change comparing models with accessibility or change in accessibility as explanatory factor, with or without control variables, and with or without removing outliers. Asterisks denote significance levels: $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***).

Dependent variable	**Change in Municipality population 2020-2023**											
Removing outliers	**None**						**IQR**					
A	1.148***	-0.413***	-0.630***	-	-	-	1.383***	0.094	0.099	-	-	-
ΔA	-	-	-	11.285***	-0.953	-1.913**	-	-	-	6.557***	0.518	1.368
D	-	74.664***	-	-	87.503***	75.524***	-	13.891	-	-	13.108	-
I	-	0.601**	-	-	0.366	-	-	0.144	-	-	0.153	-
G_i	-	-0.824	-0.614**	-	-0.174	-	-	-0.350	-	-	-0.372	-
G_e	-	0.720**	-	-	0.625*	-	-	0.137	-	-	0.145	-
G	-	-0.038***	-	-	-0.052***	-0.018***	-	-0.0023	-0.022***	-	-0.0023	-
G_{ih}	-	-3.999**	-	-	-5.091***	-	-	0.610	-	-	0.569	-
G_{eh}	-	-0.224	-	-	-0.063	-	-	-0.268	-	-	-0.277	-
G_h	-	0.124***	-	-	0.145***	-	-	-0.0012	-	-	0.0048	-
B_h	-	0.503***	0.616***	-	0.428***	0.563***	-	0.343***	0.342***	-	0.351***	0.385***
B	-	0.0025	-0.015***	-	0.013***	-	-	-0.0073	-	-	-0.0083	-0.011***
E	-	3302.396	-	-	5210.5890**	4383.453**	-	3147.896***	3768.949***	-	3101.800***	3469.928***
H	-	-3086.995*	-	-	-2069.2455	-	-	1143.617	-	-	1051.063	-
β_0	573.9***	-5899.6***	23.5	184.4*	-7178.4***	-5534.9***	94.9***	-2496.9***	-1818.0***	-46.5	-2448.2***	-1660.0***
R^2	0.27	0.89	0.87	0.38	0.88	0.86	0.16	0.61	0.58	0.05	0.61	0.58
Adj R^2	0.27	0.88	0.87	0.37	0.87	0.86	0.16	0.59	0.58	0.05	0.59	0.58
AIC	5161	4645	4672	5115	4660	4683	3015	2878	2876	3040	2878	2876
BIC	5168	4696	4690	5122	4711	4705	3021	2925	2893	3047	2925	2892
F-stat	106	166	466	173	156	358	40	24	71	12	24	71
N	290	290	290	290	290	290	208	208	208	208	208	208

Table 3: Results of model estimations of gross regional product change, comparing models with accessibility or change in accessibility as explanatory variable, with or without control variables, and with or without removing outliers. Asterisks denote significance levels: $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***).

Dependent variable	**Change in Municipality Gross Regional Product 2020-2023**											
Removing outliers	**None**						**IQR**					
A	13.035***	-1.971***	-2.319***	-	-	-	1.267*	-0.518	-0.296	-	-	-
ΔA	-	-	-	115.277***	-5.914*	-5.798*	-	-	-	35.275***	-5.407	-4.432
D	-	210.983**	-	-	274.040***	-	-	10.253	-	-	16.634	-
I	-	1.915	-	-	0.825	-	-	0.691*	-	-	0.665	-
G_i	-	-2.502	-3.754**	-	0.617	-	-	2.887**	-	-	2.924**	-
G_e	-	1.123	-	-	0.636	-	-	-0.483	-	-	-0.527	-
G	-	0.173***	0.284***	-	0.109***	0.261***	-	0.0058	0.089***	-	0.010	0.092***
G_{ih}	-	-27.317***	-13.786***	-	-32.277***	-16.536***	-	-8.738**	-	-	-8.410**	-
G_{eh}	-	2.182	-	-	2.975	-	-	0.265	-	-	0.290	-
G_h	-	0.289***	-	-	0.382***	-	-	0.202	-	-	0.169	-
B_h	-	0.258	-	-	-0.092	-	-	-0.012	-	-	-0.059	-
B	-	0.029	-	-	0.078***	-	-	0.060***	0.047***	-	0.067***	0.049***
E	-	7969.666	-	-	17168.214**	-	-	2978.015	3708.194*	-	3244.720	3561.860*
H	-	-22147.415***	-	-	-17383.921**	-	-	-7441.119***	-4807.921***	-	-6953.827***	-4773.495***
β_0	3660.0***	-13939.1**	1081.3**	-295.0	-20185.3***	448.4*	1091.5***	-2489.6	-1406.1	640.5***	-2876.8	-1321.8
R^2	0.58	0.97	0.96	0.66	0.97	0.96	0.02	0.62	0.60	0.19	0.62	0.60
Adj R^2	0.58	0.97	0.96	0.66	0.96	0.96	0.01	0.60	0.59	0.18	0.60	0.59
AIC	6184	5468	5495	6124	5486	5531	3497	3321	3318	3457	3321	3318
BIC	6191	5519	5514	6132	5537	5546	3504	3368	3338	3464	3368	3338
F-stat	401	630	1796	557	590	2098	3	25	60	48	25	61
N	290	290	290	290	290	290	208	208	208	208	208	208

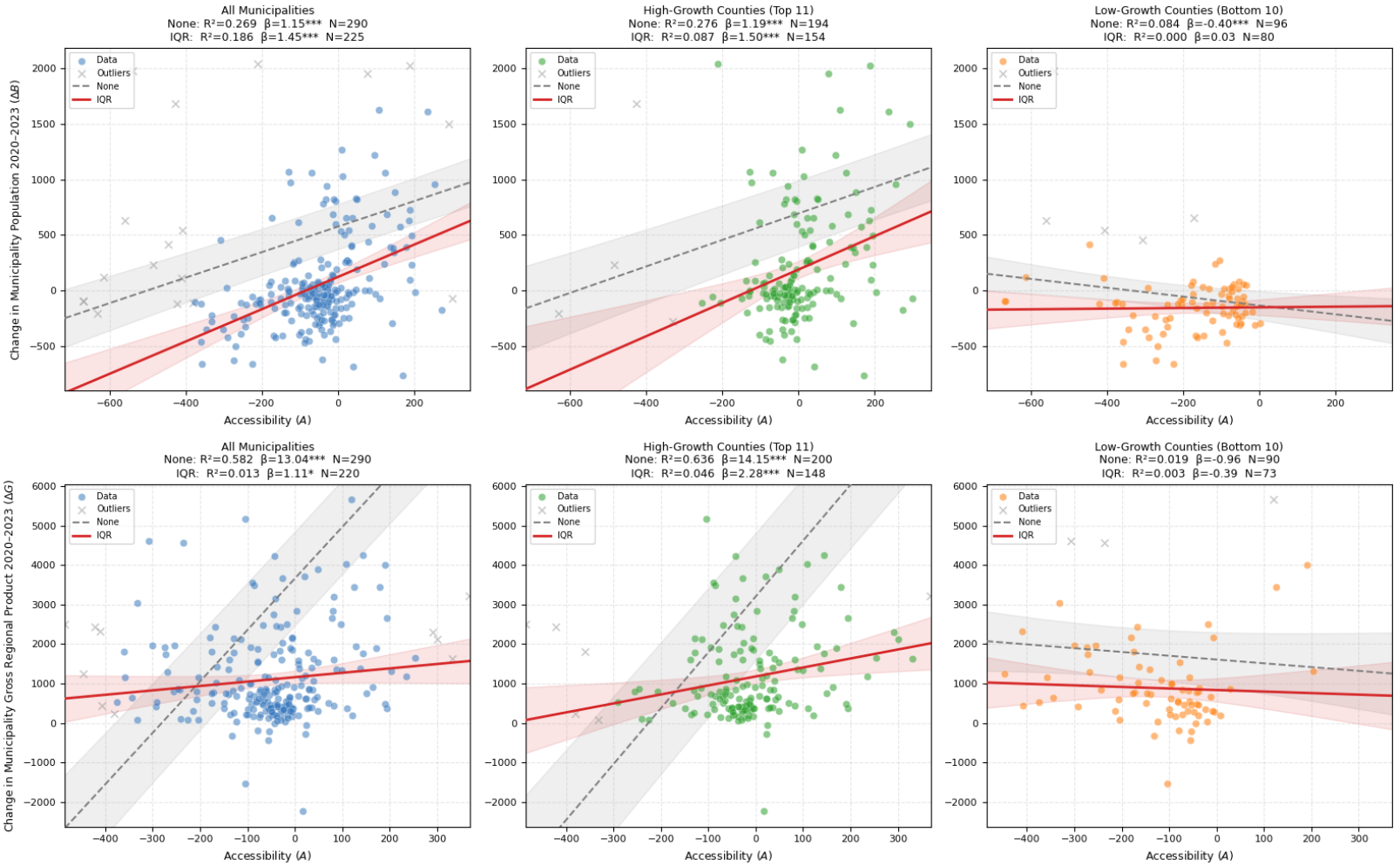


Figure 1: Model estimations of municipality population change (top) and gross regional product change (bottom) on total accessibility (A), for all municipalities, high-growth counties (top 11), and low-growth counties (bottom 10), with and without removing outliers.

4. CONCLUSIONS

This paper examines how car accessibility relates to changes in population and gross regional product across all Swedish municipalities. We estimate a range of regression models using accessibility and changes in accessibility as explanatory variables, varying the inclusion of socioeconomic controls, outlier treatment, and the split between municipalities in high-growth and low-growth counties.

While simple linear regression models show a positive relationship between accessibility and regional growth, measured by changes in population and gross regional product, suggesting that infrastructure investments inevitably lead to regional development, closer scrutiny reveals that this relationship is not robust. Adding control variables, especially regional historic development trends, turns the impact of accessibility either negative or insignificant, and removing outliers severely weakens the results. Furthermore, the simple direct relationship only holds for municipalities in high-growth counties and disappears for those in low-growth counties.

These results suggest that accessibility has not been the main driver of regional growth for Swedish municipalities between 2020 and 2023. Rather, accessibility seems to strengthen already strong regions rather than generating growth per se, implying that the effects of transport investments cannot be assumed to be uniform across regions.

Although our findings are context-specific, based on car accessibility data for Swedish municipalities during 2020–2023, a period affected by post-COVID recovery, they raise broader questions about the robustness of similar studies. Our main conclusion is that transparent robustness analyses are essential when investigating how transport accessibility relates to regional growth, as results are highly sensitive to methodological choices such as model specification and outlier treatment. Researchers using Carlino-Mills type models or other statistical approaches to analyse the relationship between accessibility and regional development must therefore demonstrate that their results are robust across model specifications.

ACKNOWLEDGEMENTS

This study is funded by the Swedish Transport Administration under Grant TRV 2024/31261.

REFERENCES

Anderstig, Christer, Svante Berglund, Anna Enström, Gunnar Isacsson, and Jonas Westin. 2021. *Markanvändning, Bostadsbyggande Och Transporter Utveckling Av Modellen Samlok-Bostad*. Manuskript. Trafikverket.

Berglund, Svante. 2025. “Effekter Av Tillgänglighetsförändringar På Omlokalisering Av Befolkning Och Bostadsinvesteringar: Ett Nytt Modellpaket.” Transportforum, Linköping, January. https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1937409&dswid=1751.

Carlino, Gerald A., and Edwin S. Mills. 1987. “The Determinants of County Growth.” *Journal of Regional Science* 27 (1): 39–54. https://doi.org/10.1111/j.1467-9787.1987.tb01143.x.

Fujita, Masahisa, Paul R. Krugman, and Anthony Venables. 1999. *The Spatial Economy: Cities, Regions, and International Trade*. MIT press. https://doi.org/10.7551/mitpress/6389.001.0001.

Hoogstra, Gerke J., Jouke Van Dijk, and Raymond J. G. M. Florax. 2017. "Do Jobs Follow People or People Follow Jobs? A Meta-Analysis of Carlino–Mills Studies." *Spatial Economic Analysis* 12 (4): 357–78. https://doi.org/10.1080/17421772.2017.1340663.

Inregia AB. 2001. *Tillgänglighet Och Lokalisering: Analyser Av Inriktningsalternativen Med Koppling Av SAMPERS Och rAps*. https://www.trafa.se/globalassets/sika/sika-rapport/sr_2001_3_u2.pdf.

Inregia AB. 2003. *Tillgänglighet Och Lokalisering: Analyser Av Inriktningsalternativ Och Infrastrukturåtgärder Med SAMLOK-Modellen*. https://www.trafa.se/globalassets/sika/sika-rapport/sr_2002_4u12_2.pdf.

Rickman, Dan S. 2010. "MODERN MACROECONOMICS AND REGIONAL ECONOMIC MODELING." *Journal of Regional Science* 50 (1): 23–41. https://doi.org/10.1111/j.1467-9787.2009.00647.x.

SCB. 2026. "Statistical Database." https://www.statistikdatabasen.scb.se/pxweb/sv/ssd/START__BE__BE0101/.

Sweco. 2025. *Samhällseffekter Av Flyget i Västernorrland*. https://www.rvn.se/contentassets/b22b2f888c244521838cc1135422e99c/flygutredning_vasternorrland_2025_huvudrapport.pdf.

Syssner, Josefina, and Albin Olausson. 2016. "Översiktsplanering i Kommuner Som Krymper." *Statsvetenskaplig Tidskrift* 118 (2): 221–45. https://www.diva-portal.org/smash/get/diva2:1057761/FULLTEXT01.pdf.

United Nations. 2025. *World Urbanization Prospects 2025: Summary of Results*. UN DESA/POP/2025/TR/NO. 12. https://population.un.org/wup/assets/Publications/undesa_pd_2025_wup2025_summary_of_results_final.pdf.

WSP. 2010. *Västerås Flygplats Regionalekonomiska Betydelse*. https://erikhjartberg.se/wp-content/uploads/2014/12/V%C3%A4ster%C3%A5s-flygplats-regionalekonomiska-betydelse-l%C3%A5nga-versionen.pdf.

WSP, and ÅF. 2014. *Flygplatserna i Västernorrlands Län - Översiktlig Analys Av Samhällsnyttan*.