Empirical Strategies and Data Requirements for Estimating Labour Market Impacts of Transport Projects*

Krista Riukula† and Touko Väänänen‡
April 22, 2025

Abstract

Transport infrastructure investment is an important tool for economic development, and its effects on labour market outcomes are often referred to when discussing the economic benefits of transport projects. However, based on theoretical models and empirical research, the effects are not straightforward. This review aims to provide researchers with a good starting point for analysing the labour market effects of transport projects. We present strategies and data requirements for empirical identification of transport projects' causal effects on labour market outcomes. We present available data sources in Finland. Furthermore, we discuss using research findings to inform policy by examining the external validity of the results, the effects of the COVID-19 pandemic on research of transport and labour market, and integrating labour market effects into cost-benefit analyses while considering the risk of double counting benefits.

Keywords: *Transport project, Labour market, Wider economic impacts, Empirical research*
JEL Classification*: R42, H43, J68, H54*

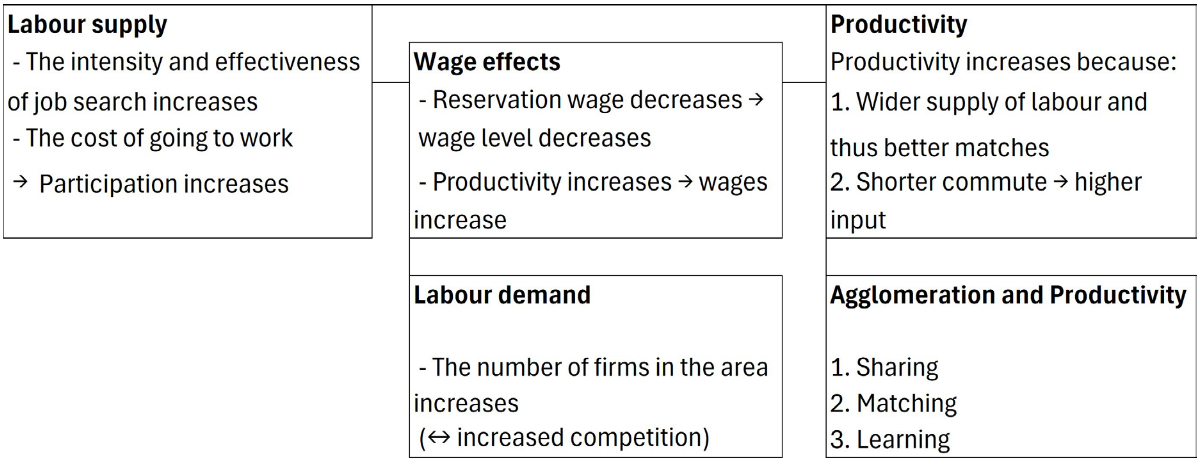
* We thank Oskari Harjunen, Antti Kauhanen, the paper's reviewers and the editor for valuable comments. This paper is partly based on Haapamäki et al. (2024) and Metsäranta et al. (2019).
† Etla Economic Research. Email: krista.riukula@etla.fi
‡ Aalto University. Email: touko.vaananen@aalto.fi

1 Introduction

Transport investment is a major part of most economies. In recent years, China has expanded its high-speed rail network rapidly. The Trans-European Transport Network plan of the European Union would extend the rail and road networks in Europe, and the core network should be completed by 2030. Cities and urban areas invest in transport to combat climate change and congestion. Additionally, the effects on labour market outcomes, including agglomeration benefits, are often referred to when discussing the benefits of transport projects[1]. Thus, the demand for well-conducted, peer-reviewed work on wider economic benefits is growing.

There are three possible channels through which transport projects may affect the labour market: 1) the supply of labour, 2) the demand for labour, and 3) matching (see, for example, Gibbons and Machin (2006) for an excellent review).[2] Figure 1 illustrates how different channels interact. For example, if labour supply increases due to lower commuting costs, the reservation wage decreases, reducing the wage level, which in turn may negatively affect labour supply. Thus, the overall impact on labour supply remains unclear and is an empirical question.

Figure 1: *The effects of lower transport costs or better accessibility on the labour market. Adapted from Metsäranta et al. (2019).*



Moreover, the agglomeration effects and the effects of other channels may overlap to some extent. Transport projects not only increase agglomeration through decreased travel times but can also influence the location of establishments. Improved transport connections attract businesses through cost savings, but on the other hand, subsequent increased competition may reduce attractiveness of these locations. Several studies have modelled the relationship between economic agglomeration and productivity. Many of these studies conclude that better transport connections lead to denser economic activity, which in turn leads to higher productivity.[3]

Duranton and Puga (2004) provide a summary of the fundamentals of urbanization and the literature focusing on these fundamentals. They identify three main channels through which economic agglomeration occurs: sharing, matching, and learning. Sharing is based on the idea that there are indivisible activities and spaces, such as factories, where broader sharing of associated fixed costs increases returns. Increased economic agglomeration also enables the sharing of a wider

[1] *Throughout this paper, we use the term transport project to describe all types of interventions and policies that alter the transport network. Where appropriate, we specify the type of project being discussed, such as a transport infrastructure project.*

[2] *This section is partly based on Metsäranta et al. (2019). Metsäranta et al. (2019) reviewed the theoretical frameworks and empirical evidence for labour market effects of transport projects. They also reviewed how these effects can be studied and incorporated into appraisal.*

[3] *Fujita and Thisse (2002) discuss economic agglomeration and the theories explaining it.*

range of intermediate goods, expertise, and risks, which in turn increases total returns. As identified in the literature, matching has two manifestations. Broader labour markets and greater availability of labour improve the probability and quality of matching between workers and jobs, and economic agglomeration can reduce delays in contract finalisation because there is a larger pool of potential contracting partners. The third channel of agglomeration benefits is learning: economic clusters, such as cities, provide favourable conditions for knowledge creation, transmission, and accumulation.

This study presents strategies and data requirements for the empirical identification of the causal effects of transport projects on labour market outcomes. We present common challenges, pitfalls, and strategies that have been used to remedy them. We also present studies that have used these strategies to study different labour market outcomes. Moreover, we highlight the data requirements for empirical work and challenges when working with spatial data. Finally, we consider some aspects relevant to transport policymaking and planning to facilitate the use of the research findings. We present the cost-benefit framework widely used in analysing transport projects and explore how to present labour market benefits and impacts in cost-benefit analyses considering the risk of double counting some benefits. We also consider the external validity of the research findings and highlight the possible challenges that the COVID-19 pandemic and increased teleworking may bring to the study of transport projects.

We contribute to the literature in four ways. First, we provide a comprehensive and concise review of reduced-form empirical strategies for causal inference and the data required to study the labour market effects of transport. Second, we provide examples of recent empirical studies that have used these empirical strategies. Third, we discuss how to include labour market effects in the cost-benefit analyses of transport projects and the risk of double counting some of the benefits. Finally, we present some challenges that increased teleworking may have in studying the labour market effects of transport projects and how they may affect the use of current estimates of the effects.

In some parts of this paper, we specifically consider Finland, as Finland's institutional and data infrastructure is familiar to us. We also highlight the Finnish datasets as they provide excellent possibilities for high-quality research. In addition, many of the specific data and policy elements are similar, at least in other Nordic countries that have administrative databases and institutions similar to Finland. Therefore, readers in other countries can benefit not only from the general but also from the more Finnish-specific sections of this review.

The remainder of this paper is organised as follows. In Section 2, we present the challenges in the empirical study of transport projects and strategies to circumvent them. We also present the data requirements, available datasets, and resources in Finland. In Section 3, we review the cost-benefit analysis of transport projects and how labour market impacts can be incorporated into cost-benefit analyses without double-counting some benefits. We also discuss the generalisability of results and offer remarks on how increased teleworking after COVID-19 may challenge empirical examination. Finally, in Section 4 we conclude by discussing future avenues for research on the labour market impacts of transport projects.

2 Empirical strategies and data requirements

We start by presenting some common challenges in inferring causal relationships between transport projects and labour market effects and strategies used to address these challenges. Some of the challenges can be circumvented with high-quality spatial and individual data, but overcoming others may require knowledge of the specifics of the project, plans, context, or other creative solutions.

2.1 Common pitfalls

The main challenges regarding empirical research on transport investment can be split into two possible sources: i) the non-random assignment of transport projects, and ii) violations of the stable unit treatment value assumption (SUTVA)[4].

[4] *We present the Rubin causal model that underlies this classification of empirical challenges in Appendix A.*

Non-random assignment can be further classified into two categories: 1) self-selection of employed and more productive workers or establishments to areas with better accessibility, and 2) reverse causality from the endogenous placement of transport investment.

Combes et al. (2008) highlight the importance of considering self-selection of individuals with high productivity into denser and more accessible areas. Their contribution highlights the need for high-quality data regarding the skills and attributes of the workers studied. Reverse causality can arise from transport projects undertaken where there is demand for them. This demand may be due to, for example, a growing population or employment. It is also possible that transport investment is targeted at economically declining regions to stimulate economic growth. In such cases, it is difficult to distinguish the impact of a transport project from a region's general trends.

Violations of SUTVA occur when transport projects induce the displacement of economic activity from areas further away to areas affected by new infrastructure. This displacement might occur on a local scale or by shifting economic activities away from other regions. If these shifts are not observed and areas that suffer from displacement are used as controls, the causal effects of transport projects will be inflated. Thus, disentangling actual growth from displacement poses a significant challenge for the assessment of the totality and causality of the labour market impacts of transport projects.

Further, we highlight a problem regarding data that affects spatial research in general: the modifiable areal unit problem (MAUP). This problem is related to the fact that spatial data are usually delivered in some kind of zonal divide, and data from different sources might not adhere to the same divide. Researchers may then have to aggregate data from different zonal divides, which can bring sizeable biases to estimations if done carelessly (Briant et al., 2010). In addition, some phenomena might appear on such a small spatial scale that studying them with large zones might not be feasible.

Next, we turn to the solutions used in the literature to solve the aforementioned problems.

2.1.1 Selection on unobservables – endogenous placement and reverse causality

The literature has dealt with the endogenous placement of transport projects mainly through three different strategies: planned route instrumental variables, historical route instrumental variables, and incidental place strategies. Additionally, minimum-cost spanning tree instrumental variables and geography based instrumental variables have been used. We briefly introduce each strategy. A more comprehensive review of the three first strategies can be found in Redding and Turner (2015).

Planned routes instrumental variables. This strategy is based on the fact that transport plans undergo many revisions, and planners usually propose different routes from which the final transport route is chosen. These planned but unrealised plans are used as instruments to eliminate the endogeneity problems caused by endogenous placement. The relevance condition for instrument validity is typically satisfied because different planned routes are usually similar in nature. The instrument validity then rests on the exogeneity of the unrealised plan. Arguments for the exogeneity of the unrealised plans to the outcome of interest are usually that the planners had different goals than labour market outcomes for the unrealised routes. Baum-Snow (2007) explains that parts of the 1947 US interstate highway system were built with military use in mind and are thus exogenous for commuting, which is the outcome of interest. Pogonyi et al. (2021) argues that in London, unrealised metro station plans are a valid counterfactual for the realised Jubilee Line Extension metro.

Historical routes instrumental variables. This instrumental variable strategy resembles the one above, but the idea is to use old transport networks as instruments for the current placement of transport routes. The argument for the exogeneity of historical routes on current labour market outcomes is that old enough transport networks (for example, Duranton and Turner (2012) use the US railroad network from the mid-1800s and Garcia-López et al. (2015) use old Roman roads) are orthogonal to current transport needs. It is argued that this orthogonality of past transport networks for current economic needs is due to structural changes in the economy. In developed countries, the economy has shifted from a more agricultural and spatially spread economy to a more knowledge-intensive and spatially concentrated economy. This structural change in the economy also means that the need for transport has changed, and that historical transport networks do not have an impact on the modern economy, except through their effect on current transport networks.

The relevance of these historical routes instrumental variables stems from the fact that transport infrastructure is less costly to build in corridors that have or had some infrastructure before. Either the geographical conditions for building are favourable, or the fact that groundwork for the infrastructure has already been done helps reduce costs for new transport infrastructure. However, some evidence suggests that historical transport networks might have long-lasting effects on economic outcomes, possibly due to agglomeration economies or zoning (Brooks and Lutz, 2019) or because they act as coordination devices for the placement of economic activity (Bleakley and Lin, 2012). The historical route strategy has been used, for example, to study the effect of high-speed railways on innovation (Hanley et al., 2022), highway effects on city growth (Duranton and Turner, 2012), and highway effects on suburbanisation (Garcia-López et al., 2015).

Incidental places strategy. In the third strategy, the identification of causal effects stems from places that, despite not being the main focus of planners, incidentally receive better transport infrastructure by virtue of being along the constructed route. The literature has identified two different ways of arguing for incidental changes in certain areas. First, researchers might search for places that are either en route to the actual areas the project is meant to connect. These types of places tend to be located along highways which are planned to facilitate movement between two cities or other major regions. Smaller villages or areas along the highway can then be considered to receive better transport infrastructure by chance (Chandra and Thompson, 2000). Fretz et al. (2022) use the construction of Swiss highway network to study income segregation and urban sprawl, combining structural modelling with similar incidental places strategy as Chandra and Thompson (2000) to estimate the parameters for the model.

Considering public transport, some areas might be treated with a station for reasons other than labour-market effects. For example, Pogonyi et al. (2021) provide evidence that some areas in London received a metro station because they were on the least-cost path between two economically motivated stations, and Tyndall (2021) argue that some areas in American cities received light rail stations by virtue of being in the vicinity of airports. However, the researcher must be wary that the incidental places do not face the same shocks as the main areas that the transport project is meant to connect. If the shocks would extend to the supposedly incidental places, this strategy would not uncover the true causal effect. The case for this strategy can also strengthened with institutional evidence that supports the assumption that the studied places received the infrastructure incidentally.

Another way to argue for incidental variation from transport projects is through a continuous accessibility measure. With sufficiently fine spatial data, researchers can accurately measure changes in accessibility that can vary even over short distances. The accessibility improvement that a transport project causes to a certain place depends on the transport network already constructed around that place. These small differences in accessibility improvements due to the previous transport infrastructure are incidental to the original plans for the placement of the transport project. Thus, this variation can be used to identify the effect of a transport project even when the project is endogenously placed. Sanchis-Guarner (2012) was among the first to provide causal evidence of transport projects' effect on individual labour market outcomes using small scale variation in accessibility. This line of argument has also been used by, for example, Börjesson et al. (2019), Gibbons et al. (2019) and Haapamäki et al. (2024) to study the effects of transport projects on worker productivity and firm-level outcomes.

Minimum-cost spanning tree instrumental variables. The fourth strategy, minimum-cost spanning tree instrumental variables, uses algorithmically constructed transport networks to instrument for a constructed transport network. The basic idea for these instrumental variables is that transport networks are sometimes explicitly planned to connect certain major cities or regions together. However, political or economic considerations may influence the exact routes and regions the roads or railways run through to connect the main regions. To instrument for this source of endogeneity, researchers have created least-cost networks to connect the plan's major cities. These hypothetical networks answer the question what the network would look like if the only objective of the plan would be to connect all cities in the plan with minimal construction costs. The minimum cost networks can be based on, for example, terrain cover and elevation data that affect the construction costs of transport infrastructure. Researchers have also used straight line networks as instruments. The exclusion restriction of these instruments is that the economic development of areas that the minimum cost network passes through are only affected by the construction of the actual network.

This strategy has been used to study, for example, highways' effect on organisation and efficiency of manufacturing production (Ghani et al., 2015), local GDP growth in peripheral regions (Faber, 2014), sectoral employment in China (Dong, 2018), and transport network's effect on GDP per capita (Banerjee et al., 2020).

Geography based instrumental variables. The fifth instrumental variables strategy is based on geographical features of areas that affect the construction costs of transport infrastructure. The exclusion restriction here is that the geography of the areas affects the economic outcome only through its effect on transport infrastructure construction. The geography of areas can also be a source of productivity gains and similarly to the historical routes instrumental variable, it is usually argued that the geographic features used do not matter for modern economy as in, for example Holl (2016), who argues that aquifers are not meaningful for modern day manufacturing productivity. These instruments are often used in addition to other instrumental variable techniques. Zhang et al. (2020) use the average slope of a county with historical postal routes, Martín-Barroso et al. (2020) use terrain ruggedness and altitude with historical population density and market access to instrument for current day accessibility, while Faber (2014) uses the geographical features of terrain to construct the least-cost routes between major Chinese cities as a part of a minimum-cost spanning tree instrumental variable strategy.

Other sources for exogenous variation. Other plausible exogenous variations in transport networks have also been used in the literature. These research settings make use of the differences-in-differences method, and thus rely on the validity of the assumptions of the method. The difference-in-differences method utilises natural or quasi-experimental situations with panel data based on the idea that the selection into the intervention is driven by a time-invariant unobservable factor. The effect of the selection process on outcomes can be eliminated by comparing the changes in outcomes between experimental and control groups (see, for example, Pekkarinen, 2006). This enables the establishment of a cause-and-effect relationship rather than a mere correlation, and produces valid estimates of the effects of individual interventions.

Åslund et al. (2017) use difference-in-differences matching estimator to create observationally similar treatment and control groups. Although matching estimators are usually used to remedy selection bias, Åslund et al. (2017) specifically state that the use of a matching estimator balances the observed qualities of the treatment and control groups. Feng et al. (2023) use propensity score matching and spatial difference-in-differences model to study the effect of high-speed rail to the mobility of high-skilled labour in China. Tyndall (2017) exploit a natural disaster that caused the destruction of parts of the transport network as exogenous variation to identify the effect of transport infrastructure on employment. He uses the flooding of a metro tunnel in New York caused by Hurricane Sandy in 2012 to estimate the effect of metro access on the employment of people in the vicinity of the flooded metro line. Bütikofer et al. (2022) examine the impact of access to larger labour markets on wages and employment by exploiting the opening of the Öresund bridge between Denmark and Sweden using a difference-in-differences estimator. To strengthen their case, they use extensive pre-treatment analyses and robustness checks to convince the reader of the validity of their identification.

Nilsson (1991) found that *ex ante* analysed rate of return did not affect which transport projects were selected to the Swedish 10 year transport plan. In addition, Eliasson et al. (2015) found that the results of cost-benefit analyses did not affect the selection of transport projects in Norway and Sweden. Even if civil servants' choices for transport projects were affected by the cost-benefit analysis (CBA) results, the choices made by politicians were only mildly affected. Börjesson et al. (2019) use these results as one piece of evidence to argue that the placement of transport infrastructure is actually exogenous for the economic phenomena examined, alleviating possible problems with assignment on unobservables. This line of thought, of course, depends on the available research on the effect of CBA's or other *ex ante* metrics on choices of transport projects. The results from Norway and Sweden do not mean that the choices of politicians in other countries would not be influenced by *ex ante* economic evaluations of transport projects.

Strengths and weaknesses. Each of the presented strategies have their own strengths and weaknesses. The planned routes, historical routes, and incidental place strategies typically require further institutional arguments to support claims that they satisfy the exclusion restriction. The least cost spanning tree strategy on the other hand requires knowledge of the major cities of regions that the planners were targeting to connect. Additionally, the amount of work required to obtain suitable data for each strategy can be substantial. A historical route strategy may require digitising historical maps, and a planned route strategy may require finding and digitising old transport plans. Using a continuous accessibility measure to find small differences in accessibility, which planners did not have in mind, requires data to calculate accessibility on a sufficiently small spatial scale. However, if the institutional and data requirements for one of these strategies are met, they are typically considered solid ways to estimate the causal effects of transport projects.

Natural disasters that dismantle transport networks are prime candidates for random variation in transport networks, but the time frames from the destruction of a part of a network and its rebuilding may be too short to identify meaningful long-run effects. It is also possible that people's responses to building and dismantling transport infrastructure differ (for example, Gibbons et al. (2024) found that the decrease in population after train station closures was smaller than the increase in population after their opening).

2.1.2 Violations of SUTVA – displacement effects

Violations of SUTVA are an important concern in many research questions regarding the causal effects of transport projects. Changes in labour market outcomes due to transport projects may be area-specific and come at the expense of areas that do not receive investment. This may reduce the total effect of transport investment and even make it negligible (Chandra and Thompson, 2000; Holl, 2004; Pogonyi et al., 2021). Studying the possible displacement to assess the total economic effects of transport investment requires data not only from the vicinity of the investment but also from a larger area that might be negatively affected by the relocation of jobs and establishments.

Redding and Turner (2015) suggest a reduced-form approach with three types of areas to assess the displacement of economic activity. One of the areas is where the investment is located (i), another is where activity might be shifted away from (ii), and a third is not affected by investment (iii). By comparing areas (ii) and (iii), we find the amount of displaced economic activity d. Then, by comparing areas (i) and (ii), we can estimate the amount of shifted and added economic activity, $2d + a$. By combining these estimates, we can then infer the extent to which the transport project increased economic activity and how much of it merely shifted from other regions. Naturally, this identification relies on assumptions that may not be met in practice. Redding and Turner (2015) discuss those briefly.

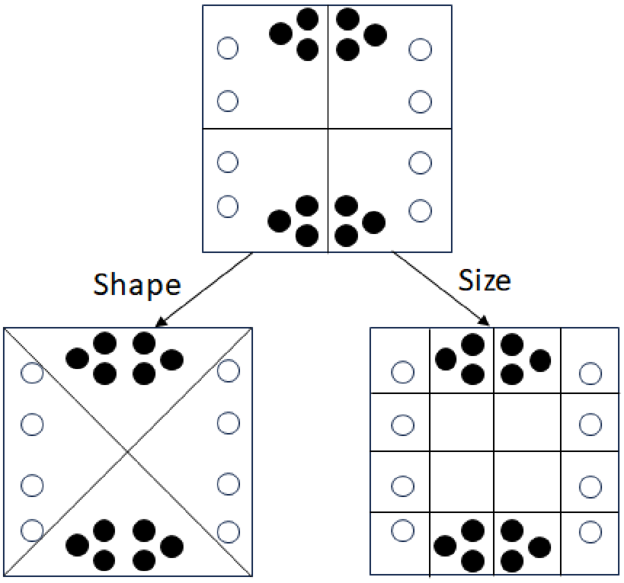
2.1.3 Modifiable areal unit problem

Gehlke and Biehl (1934) were one of the first to note that zonal divide can affect correlation coefficients and other metrics when these metrics are calculated with different divides. This modifiable areal unit problem (MAUP) stems from two properties of dividing geography into smaller units: (i) the level of aggregation or the size of the zone and (ii) the way boundaries are drawn between the zones or the shape of the zone.

We illustrate these effects with an example adapted from Briant et al. (2010). In Figure 2, imagine that the black dots describe high-productivity workers and the white dots describe low-productivity workers. First, each zone has the same density and number of high-productivity and low-productivity workers, so density and productivity do not seem to be associated. The size effect is illustrated by densifying the zonal divide, which leads to variation in the density of workers in each zone such that higher density zones have all the high-productivity workers, and lower density zones are left with the low-productivity workers. The shape effect is illustrated by modifying how the original four zones are drawn. This similarly leads to lower density zones with low-productivity workers and high-density zones with high-productivity workers, but to a lesser extent.

This problem is closely related to gerrymandering, in which spatial units are intentionally reshaped to provide an outcome that is beneficial for the actor. The example above shows how someone with the possibility of choosing the spatial divide used in an analysis could influence the results. Thus, understanding these effects can help, for example, in assessing the plausibility of results from research settings that involve spatial data.

Figure 2: *Illustration of the modifiable areal unit problem. Adapted from Briant et al. (2010).*



Briant et al. (2010) study how the size and shape effects distort the analysis of spatial concentration, agglomeration economies and trade determinants. They conclude, based on simulations and French administrative data, that equally shaped and smaller spatial units create smallest distortions to the data. They also note, that the size of the units is more important than their shape. Tveter et al. (2022) examine the error in transport costs when smaller spatial units are combined with larger units for analysis. Using simulated data, they show that the error in transport costs is minimised when smaller units are combined with larger ones through the harmonic mean.

Suggestions from the MAUP literature can be summarised into three key points: (1) data should be spatially disaggregated and the spatial units should be small and uniformly shaped; (2) when combining spatial units for analysis, one should use the harmonic mean for transport costs; and (3) the distortions from MAUP are of second order when compared to problems from wrong specification of empirical models.

2.2 Data requirements and resources

2.2.1 Individual- and firm-level data

The data requirements for evaluating the effects of transport projects on labour market outcomes are substantial.[5] First, individual-level panel data are necessary to enable a thorough evaluation of the effects and account for factors such as changes in migration behaviour. For example, Heres et al. (2014) showed that transport projects can result in selective migration, and failing to consider this can lead to biased conclusions about the project's effects. Additionally, individual- and firm-level panel data offer an opportunity for a broader examination of both mechanisms and effects, as demonstrated by Gibbons et al. (2019).

[5] *This section is partly based on Metsäranta et al. (2019).*

Second, for accurate assessments, precise information about residential or business locations is required, for example, at the 250 × 250 m grid level or even at the building level. If the location is at a coarse level, such as a municipality, much of the variation in travel times or accessibility is masked under the municipal-level measurement. This leads to measurement error in the independent variables and attenuation bias. More detailed data, such as those at the grid level, allow for more accurate measurement of accessibility and less biased estimates.

Third, data spanning a long period of time are required. When comparing different areas or cities, it is important to demonstrate that the research design is credible, and that the areas have followed the same trends over time, such as employment, for an extended period. Additionally, areas need to be monitored for years after the project to capture all potential effects. Data covering a broader area also enable the assessment of displacement effects.

Statistics Finland has datasets that satisfy these requirements. The FOLK modules (formerly FLEED) contain individual-level panel data, providing comprehensive information on skills (such as education) and employment (such as workplace, annual earnings, and income). These data enable us to control for both the observed and unobserved characteristics of workers. For example, data on firms are available from Statistics Finland's financial statement statistics, and municipal workplace statistics include information such as the industry of the workplace, turnover, and number of personnel. Using workplace-level data enables the use of a more comprehensive productivity measure and controls for observed and unobserved productivity differences at the workplace level (see, for example, Gaubert, 2018). Additionally, it is possible to obtain location data for both the individual's place of residence and workplace at the grid level (250 × 250 m), and in some cases, even at the building level, from Statistics Finland.

2.2.2 Accessibility data

Transport networks and their changes can be measured in various ways. The number of public transport stops in an area, the length of the road network in an area, or changes in them are some simpler ways to describe the transport infrastructure in an area. A list of transportation related data sources for research in the Finnish context and detailed information about them can be found in Appendix B. Much of the data can be found for the entire country, but more sophisticated tools, such as travel demand models, are currently only available for certain areas.

However, the simple measures neglect important features of the transport infrastructure. First, transport networks are used by people to reach activities that they wish to do (Jones, 1979). Therefore, to fully assess the importance of a change in a transport network, researchers must measure how the change affects the ease of reaching different activities. Second, measures that concentrate only on transport infrastructure cannot be used to study policy interventions, such as congestion charging or decreases in speed limits. Third, measures that concentrate on one travel mode neglect mode choice, which may be an important determinant of the impact on travel. For example, the public transport network has economies of scale for travellers, where every additional traveller makes it more profitable to add services, which then improves the travel times of everyone using public transport.

Measuring accessibility. Accessibility, or market potential, measures provide a unified framework for assessing the effects of different types of transport projects. These accessibility measures take the following forms:

$$A_i = \sum_j O_j f(\tau_{ij}), \tag{1}$$

where i and j indicate the areas, A_i is the accessibility of area i, O_j is a size measure that indicates the importance of the area, such as the number of workers, number of residents, or wage sum. Finally, τ_{ij} is a measure of transport cost, such as travel time or the generalised cost of travel, which combines the pecuniary cost of travel and travel time to a single measure of travel cost. The importance of an area j to the accessibility of i is mediated by the decay function $f(\tau_{ij})$. This function is typically chosen to be the inverse of the travel cost $f(\tau_{ij}) = \tau^{-\alpha}$ or the exponential function $f(\tau_{ij}) = \exp(-\beta\tau_{ij})$. The parameters α and β control how quickly the importance of places diminishes as a function of the transport cost.

Researchers have adopted different approaches to select these parameters. A typical assumption is that the parameter for the inverse function is $\alpha = 1$ (for example Gibbons et al., 2019). Other authors have estimated these parameters with gravity models using travel diaries or commuting flow data, as the accessibility measure has a gravity-model-like appearance. The parameter is then interpreted as the coefficient of the travel cost in the gravity model (see, for example, Graham and Melo (2011) for the inverse function and Lee (2021) for the exponential function). Knudsen et al. (2022) calibrate the parameter in their exponential decay function by maximizing the coefficient of determination of their wage equation.

An important detail in measuring accessibility is the transport cost τ_{ij}. In some applications where modal choices do not play a large role, the researcher may be able to use travel times from a certain dominant mode, usually the car, as the travel cost between the areas. These situations can arise when the research considers inter-city travel or other national travel, where the mode share of car travel can exceed 90 % as the length of the journey increases. In other cases, the inclusion of alternative modes of transport, including travel times and pecuniary costs, is crucial. This is especially true in cities that have sizeable mode shares of public transport, cycling, and walking and when the transport project of interest concerns modes other than the dominant one. The importance of accounting for congestion in travel times was noted in Graham (2007b) and Graham and Dender (2011).

Most studies use the travel cost of a single mode as the travel cost between areas (Gibbons et al., 2019; Graham, 2007b; Knudsen et al., 2022); however, some have included other modes in the transport cost (for example, Börjesson et al., 2019). Combining different travel modes into a single transport cost measure is achieved by mode share weighted sum of all travel costs. For example, in Börjesson et al. (2019) the generalised cost for each mode was first calculated considering travel time and pecuniary travel costs. Travel time of each mode was converted to monetary terms with a mode specific value of travel time savings. These generalised travel costs are then aggregated into a single cost measure by weighting the generalised cost of each mode by its mode share and summing these weighted costs together.

Another suggested way of aggregating multiple modes into a single transport cost measure is to use the inclusive value from a logit mode choice model as the cost measure. In the context of random utility models, this value is the expected maximum utility an agent receives from a choice (Cascetta, 2009). These types of measures are typically present in structural models of transport (for example, Allen and Arkolakis, 2022; Tyndall, 2021) but are not extensively used in reduced-form work.

3 Remarks on transport policy

Research on the labour market impacts of transport projects is used to provide information about the effects of transport projects and policies. In policymaking and planning, the focus is typically on unrealised projects that are being planned or approved for construction or further planning. To this end, the external validity and generalisability of previous research are of crucial importance, as using excessively high or low estimates for effects may lead to over- or under-investment in transport infrastructure. For example, impact assessments based on excessively high agglomeration elasticities of accessibility may overstate the productivity gains from agglomeration.

Effective methods for impact assessment also help in realising the objectives of the investment, which is not always easy. Welde and Tveter (2022) investigated the impact of ten transport projects on commuter traffic, population, employment, and the number of new businesses using synthetic control groups in Norway. Significant time savings (5–60 min) in travel time were calculated for all projects, except for one. Apart from one transport project, none of the projects had a positive impact on all the four aforementioned factors; some had negative effects, while others had positive effects. Overall, the projects did not have a significant impact on the listed factors, although these factors were included in the project objectives.

3.1 Generalisability of the results

For policymaking and planning, the external validity of the results is crucial. Several empirical and theoretical studies highlight how difficult it is to generalise the labour market impacts of a specific transport project to other projects.[6] The literature has found varying effects depending on factors, such as the functional form of structural models or empirical perspectives. In general, the literature highlights the importance of considering transport projects conducted in similar institutional and transport contexts. Studies that have aimed for estimates to be used in, for example, cost-benefit analyses have typically used accessibility changes by transport investment to help generalise their effect (for example, Börjesson et al., 2019; Knudsen et al., 2022; Haapamäki et al., 2024).

According to Oosterhaven and Knaap (2017), the relationship between production and infrastructure found in historical macro data cannot be directly applied to the evaluation of new projects; instead, several factors, such as the characteristics of the transport project and its location, must be considered. Venables et al. (2014) also emphasize the importance of conducting project-specific assessments. According to Deng's (2013) review, three main factors underlie the differences in impacts that are considered in project evaluation: 1) country-specific characteristics (such as the ability to efficiently utilise new transport projects), 2) the type and quality of the transport project, and 3) the condition of the existing transport network.

Moreover, relative location of the project may have an impact on labour market outcomes. Koster et al. (2022) utilize planned and implemented railway projects mainly in Japan to examine winning and losing regions. Based on both theoretical models and empirical research, they demonstrated how a railway connection may not necessarily boost employment in remote or mid-range areas and could even decrease it. Their findings highlight the importance of relative location within the railway and road network and help explain why some regions lose, while others gain.

3.2 Cost-benefit analysis

The main economic analysis framework for assessing transport projects is cost- benefit analysis, which assesses the social profitability of a transport project through its effects on the transport market. As we have seen, transport projects have effects beyond the transport market in labour markets. In a context where cost-benefit analyses are used to analyse the rationale for projects, it is vital to understand the relationship between benefits that are assessed in the cost-benefit analyses and the outcomes that occur in labour markets due to assessed projects. As commuters trade off commuting time and wages, calculating the benefits from increased employment and wages can lead to double counting of benefits if proper care is not exercised. Nevertheless, even if the social value of labour market benefits is already evaluated in the cost-benefit analyses, estimates of the different labour market outcomes can be of interest to policymakers and planners. Labour market analyses of transport projects can be used to understand how the projects might help with labour market objectives the politicians have.

The basic premise of cost-benefit analysis is to evaluate the project based on its consequences with the theoretical basis drawn from welfare economics (Boadway et al., 2006). Despite its widespread use, the use of cost-benefit analysis has also been debated, partly because of the many ethical and theoretical decisions that are needed to arrive at such a framework. More complete treatments of cost-benefit analysis can be found from Boadway et al. (2006) and Drèze and Stern (1987), and summaries of the criticisms around the method can be found from, for example, Næss (2006) and Van Wee (2012). Here, we concentrate on the practical use of cost-benefit analysis to evaluate transport infrastructure investments.

Typically, cost-benefit analysis of transport infrastructure investments is performed in a partial equilibrium setting, where only impacts that happen in transport markets are analysed. A model is used to compare a world with the analysed transport infrastructure investment in a world without. Transport infrastructure is durable, so the effects of these projects have been appraised for many years. A typical period ranges from 30 to 50 years (Sartori et al., 2014). Estimates of yearly

[6] *This subsection is partly based on Metsäranta et al. (2019).*

benefits are usually assessed by modelling the start and end years of the assessment period and interpolating the benefits in the years between. The benefit calculations can be made more accurate by additionally modelling some years between the start and end years.

Generally, the impacts that are appraised in the process are the effects on time savings, traffic safety, local pollution and noise, greenhouse gases, ticket revenue and operating costs of public transport, and maintenance costs of infrastructure (Sartori et al., 2014). The effects are monetised with unit values that can be derived from, for example, revealed or stated preference experiments, mitigation or damage costs, costs of meeting targets, or hedonic prices (Koopmans and Mouter, 2020). The monetised benefits of completing the project are then compared to the costs of investment. If the benefits exceed the costs, the project is deemed socially beneficial. This rather simple cost-benefit test is derived from the Kaldor-Hicks principle. The principle states that if beneficiaries of a resource reallocation could compensate those who suffer in a way that makes the reallocation a Pareto improvement, the reallocation is socially beneficial.

3.3 Labour market impacts and the risk of double counting benefits

Increased economic activity and labour market effects are natural candidates for the evaluation of benefits in a cost-benefit analysis of transport projects. Transport projects might decrease unemployment if reduced commuting time encourages workers to extend their job search range, accept work from employers further away (Manning and Petrongolo, 2017), or increase wages if productivity rises due to agglomeration effects (Redding and Turner, 2015). Nevertheless, researchers need to be careful about the benefits they use in conjunction with each other, as changes in the transport market capitalise in other markets, such as real estate and labour markets. Thus, carelessly adding different benefits together can lead to double counting some benefits.

Considering labour market benefits, it's important to note that models of commuting and employment posit that people trade off commuting times and wages on the margin (for example, Eliasson and Fosgerau, 2019; Manning and Petrongolo, 2017). Thus, the value of reduced commuting time already measures the welfare effects of increased employment through transport investment and including benefits from increased employment would lead to a double counting of these benefits.

However, all benefits from decreased travel time are assessed through the value of travel time savings only in perfect and competitive markets. The connection between transport and labour markets is imperfect. Venables (2007) shows that decreased travel time can lead to increased productivity and tax revenue through increased agglomeration. These effects are external to people's commuting decisions and thus not considered in their value of travel time savings. This result sparked literature that has attempted to estimate these external agglomeration benefits to incorporate them in cost-benefit analyses (Börjesson et al., 2019; Graham, 2007a,b; Knudsen et al., 2022). In the studies productivity is measured mainly using wages. The underlying assumption is that in market equilibrium, wages reflect productivity[7].

Unfortunately, incorporating these results to national guidelines has been slow, as the risk of double counting exists even with the aforementioned estimates. Duranton and Puga (2004) divide agglomeration benefits emanating from three different channels: (i) matching, (ii) sharing and (iii) learning.[8] The difficulty of incorporating agglomeration benefits into cost-benefit analyses stems from the fact that workers take the effect of matching into account when making commuting choices. The underlying reason is similar as in the case of increased employment. Commuters trade off commuting time and wages, and thus consider their possibilities to earn increased wages when choosing jobs in further away locations which shows in their value of travel time savings for commuting trips (Eliasson and Fosgerau, 2019). Additionally, firms

[7] *Empirical evidence shows a clear connection between productivity and wages, although the relationship may not be direct 1-to-1. For specific studies see, for example, Hellerstein et al. (1999); Lazear et al. (2022)*

[8] *Matching referring to more productive matches between firms and workers, sharing to the possibility of sharing indivisible facilities or intermediate suppliers, and learning to the dissemination of knowledge and best practices.*

most likely experience the benefits of sharing through, for example, smaller logistics costs that are represented by travel time savings of vans and trucks in cost-benefit analysis. However, commuters do not consider the learning others will gain from interacting with them. Therefore, learning benefits are left as external for the choices of commuters.

Separating different mechanisms of agglomeration has been under research, but has mostly concentrated on showing the relevance of each mechanism to the benefits of agglomeration (Combes and Gobillon, 2015). Recent efforts have produced estimates that leave matching effects out by concentrating on accessibility between workplaces (Börjesson et al., 2019; Knudsen et al., 2022; Haapamäki et al., 2024). Benefits assessed with these estimates can plausibly be added to cost-benefit analyses that do not account for travel between workplaces. Nevertheless, the question remains how to incorporate agglomeration benefits into cost-benefit analyses in settings where travel time savings for travel between workplaces are assessed. Hence, work to include agglomeration benefits in cost-benefit analyses without double counting is still underway. However, the tax wedge component of productivity gains from transportation infrastructure improvements can be added to the analyses (Venables, 2007).

3.4 Effects of the COVID-19 pandemic and teleworking

The COVID-19 pandemic and subsequent increase in teleworking may affect the labour market effects of transport projects. The effects are likely to come through changes in labour markets as pandemic's effects on travel are likely to be small based on historical evidence (Eliasson, 2022). Increase in teleworking seems to be a lasting consequence, as several studies have documented favourable experiences for employees (for example, Barrero et al., 2023; Beck et al., 2020; Barrero et al., 2021). Based on these favourable experiences, researchers have predicted a sustained increase in teleworking as especially workers who find teleworking productive are willing to continue it (Mohammadi et al., 2023).

Barrero et al. (2023) is especially illuminating in the possible labour market consequences of increased teleworking. They consider the effects of teleworking on wages, productivity, and innovation in light of theory and available empirical evidence, as well as their own survey that has been conducted monthly since 2020. They conclude that in jobs that can be done remotely, teleworking increases labour supply and work amenities. This in turn puts downward pressure on wages in these jobs. The results concerning productivity are mixed. They offer several remarks, which suggest that the effect of teleworking on productivity varies between tasks, workers, whether we are concerned with short- or long-term effects, and whether teleworking is full-time or part-time. Specifically, teleworking for one or two days seems to increase productivity, and senior employees are able to increase their productivity by teleworking. However, the increased productivity comes at the expense of not being able to direct junior employees, decreasing junior employees rate of learning. The empirics on this topic are still developing however, and the researchers note that there is still much room for studying the effects of teleworking on labour market outcomes.

How does increased teleworking then affect the empirical study of transport infrastructure's labour market outcomes? Certain answers are hard to give as studies that have been published after the pandemic mostly study the effects of transport infrastructure before the pandemic occurred. The basic mechanism that links labour market outcomes to transport infrastructure is a decrease in transport costs and a subsequent increase in the accessibility of jobs and employees. As teleworking decreases commuting, improvements in commuting times may have smaller impacts on labour productivity driven by better employer-employee matches. Recent studies also highlight different commuting costs for men and women; thus, increased teleworking may lead to more equitable employment and wage outcomes in the future (Le Barbanchon et al., 2020).

Setting aside labour market outcomes, teleworking might challenge the measurement of workplace accessibility. Workplace accessibility uses the number of workplaces in a given location as the size measure in Equation 1. Especially in studies of agglomeration, this measure is used to represent access to other workers which is thought to facilitate agglomeration effects. However, if a sizeable proportion of workers telework, effective access to other workers is clearly diminished and accessibility is measured with error. If the amount of teleworking cannot be controlled for, the estimates of the agglomeration elasticity of productivity are probably diminished. In addition, using former estimates of agglomeration elasticity can lead to overestimated agglomeration benefits if these benefits are tied to effective workplace accessibility.

This can be illustrated by the results from Delventhal et al. (2022) who used a structural model in which teleworkers could be set to affect agglomeration benefits to study the effects of teleworking on different outcomes in Los Angeles. They compare an increase in teleworking from around 3 % to 33 % and examine how having teleworkers contribute to agglomeration benefits affects different outcomes. They find that when increase in teleworking does not affect agglomeration benefits, increases in wages are significantly larger compared to a situation when increased teleworking decreases agglomeration benefits. Based on these results, having teleworkers affect agglomeration (measuring workplace accessibility without considering teleworkers) clearly inflates the labour market benefits of increased teleworking.

4 Discussion

We have reviewed challenges in estimating causal effects of transport projects on labour market outcomes and strategies used to address them. We also provided examples of studies that have used these strategies to study labour market outcomes. The general message based on empirical studies is that each project has unique effects on labour markets. In developed countries with mature transport systems, the effect of a single transport project is typically incremental. In addition, transport projects that seemingly improve the transport system can still have negative consequences for labour market outcomes such as city employment through general equilibrium effects.

For decision-making and planning purposes, the uniqueness and ambiguous direction of the impacts of a single transport project highlight the need for an accurate understanding of the underlying mechanisms that connect labour markets and transport markets. The *ex ante* assessment of these impacts also requires good tools and methods so that transport projects can be designed in a way that they reach their intended goals. In addition, advances are required to incorporate labour market benefits that are external to travellers in the cost-benefit analyses of transport projects.

Existing research on the effects of transport projects on labour markets mostly considers settings in which new transport infrastructure is built, although some related studies consider the loss of access due to political restricions (Ahlfeldt et al., 2015; Redding and Sturm, 2008). Transport network failures due to, for example, natural catastrophes have been used more in examining transport outcomes (for example, Zhu et al., 2010; Xie and Levinson, 2011) than labour market effects (for example, Tyndall, 2017) and could be an interesting avenue for research on labour market impacts. However, the temporary nature of these disruptions might provide too short a period for changes to occur. A similar avenue would be to study conscious decisions to dismantle some transport infrastructure by, for example, shutting down rail stations or other downgrades to public transport routes. However, compared to natural transport network failures, these settings may suffer from endogeneity problems similar to the construction of a new transport infrastructure. Moreover, research efforts could be targeted towards how labour market gains from transport improvements are distributed across workers and how they impact income inequality and the gender wage gap following the work by, for example, Bütikofer et al. (2022) and Le Barbanchon et al. (2020).

Finland's excellent data provides possibilities for high-quality research on labour market effects of transport projects. Accurate individual-level data on labour market outcomes are available from Statistics Finland. Cities, public transport authorities, and governmental institutions provide high-quality data on Finland's transport networks. However, the lack of data on historical transport networks poses a challenge to some examinations. Although records of older transport networks exist, the earliest digitised open-source versions tend to be from the mid-2010s. Most of the data provided concern the current state of the transport network, with little information about its state in previous years. Thus, constructing historical transport data may require substantial work to reconstruct the transport networks from previous years.

Recent urban transport infrastructure investments that would provide interesting quasi-experimental settings include local train (Ring rail line), metro (West metro), and tram (Tram line 15) investments in the capital region of Finland, and a tram investment in Tampere. Examples of regional transport infrastructure investments that could warrant investigation include the Lahti rail bypass that was completed in 2008 and upgrading the European road E18 to a motorway. There are also plans for large rail investments between Helsinki and Turku, as well as Helsinki and Tampere, which could benefit from the *ex ante* assessment of labour market effects.

References

Ahlfeldt, G. M., S. J. Redding, D. M. Sturm, and N. Wolf (2015). The economics of density: Evidence from the Berlin Wall. *Econometrica 83* (6), 2127–2189.

Allen, T. and C. Arkolakis (2022). The welfare effects of transportation infrastructure improvements. *The Review of Economic Studies 89* (6), 2911–2957.

Åslund, O., I. Blind, and M. Dahlberg (2017). All aboard? commuter train access and labor market outcomes. *Regional Science and Urban Economics 67*, 90–107.

Banerjee, A., E. Duflo, and N. Qian (2020, June). On the road: Access to transportation infrastructure and economic growth in China. *Journal of Development Economics 145*, 102442.

Barrero, J. M., N. Bloom, and S. J. Davis (2021, April). Why Working from Home Will Stick. (28731).

Barrero, J. M., N. Bloom, and S. J. Davis (2023, November). The Evolution of Work from Home. *Journal of Economic Perspectives 37* (4), 23–49.

Baum-Snow, N. (2007). Did highways cause suburbanization? *The quarterly journal of economics 122* (2), 775–805.

Beck, M. J., D. A. Hensher, and E. Wei (2020, October). Slowly coming out of COVID-19 restrictions in Australia: Implications for working from home and commuting trips by car and public transport. *Journal of Transport Geography 88*, 102846.

Bleakley, H. and J. Lin (2012). Portage and path dependence. *The quarterly journal of economics 127* (2), 587–644.

Boadway, R. et al. (2006). Principles of cost-benefit analysis. *Public Policy Review 2* (1), 1–44.

Börjesson, M., G. Isacsson, M. Andersson, and C. Anderstig (2019). Agglomeration, productivity and the role of transport system improvements. *Economics of Transportation 18*, 27–39.

Briant, A., P.-P. Combes, and M. Lafourcade (2010). Dots to boxes: Do the size and shape of spatial units jeopardize economic geography estimations? *Journal of Urban Economics 67* (3), 287–302.

Brooks, L. and B. Lutz (2019). Vestiges of transit: Urban persistence at a microscale. *Review of Economics and Statistics 101* (3), 385–399.

Bütikofer, A., K. V. Løken, and A. Willén (2022). Building bridges and widening gaps. *Review of Economics and Statistics*, 1–44.

Callaway, B., A. Goodman-Bacon, and P. H. C. Sant'Anna (2024, February). Difference-in-differences with a Continuous Treatment. Working Paper 32117, National Bureau of Economic Research.

Cascetta, E. (2009). *Transportation systems analysis: models and applications*, Volume 29. Springer Science & Business Media.

Chandra, A. and E. Thompson (2000). Does public infrastructure affect economic activity? Evidence from the rural interstate highway system. *Regional Science and Urban Economics 30*, 457–490.

Combes, P.-P., G. Duranton, and L. Gobillon (2008). Spatial wage disparities: Sorting matters! *Journal of urban economics 63* (2), 723–742.

Combes, P.-P. and L. Gobillon (2015). The empirics of agglomeration economies. In *Handbook of regional and urban economics*, Volume 5, pp. 247–348. Elsevier.

Cox, D. R. (1958). Planning of experiments.

Delventhal, M. J., E. Kwon, and A. Parkhomenko (2022, January). JUE Insight: How do cities change when we work from home? *Journal of Urban Economics 127*, 103331.

Deng, T. (2013). Impacts of transport infrastructure on productivity and economic growth: Recent advances and research challenges. *Transport Reviews 33* (6), 686–699.

Digitransit (2024). Digitransit. https://digitransit.fi/en/. Accessed: 2024-05-02.

Dong, X. (2018, October). High-speed railway and urban sectoral employment in China. *Transportation Research Part A: Policy and Practice 116*, 603– 621.

Drèze, J. and N. Stern (1987). The theory of cost-benefit analysis. In *Handbook of public economics*, Volume 2, pp. 909–989. Elsevier.

Duranton, G. and D. Puga (2004). Micro-foundations of urban agglomeration economies. In *Handbook of regional and urban economics*, Volume 4, pp. 2063–2117. Elsevier.

Duranton, G. and M. A. Turner (2012). Urban growth and transportation. *Review of Economic Studies 79* (4), 1407–1440.

D'Haultfœuille, X., S. Hoderlein, and Y. Sasaki (2023, June). Nonparametric difference-in-differences in repeated cross-sections with continuous treatments. *Journal of Econometrics 234* (2), 664–690.

Eliasson, J. (2022, March). Will we travel less after the pandemic? *Transportation Research Interdisciplinary Perspectives 13*, 100509.

Eliasson, J., M. Börjesson, J. Odeck, and M. Welde (2015). Does benefit–cost efficiency influence transport investment decisions? *Journal of Transport Economics and Policy (JTEP) 49* (3), 377–396.

Eliasson, J. and M. Fosgerau (2019). Cost-benefit analysis of transport improvements in the presence of spillovers, matching and an income tax. *Economics of Transportation 18*, 1–9.

Faber, B. (2014, July). Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System. *The Review of Economic Studies 81* (3), 1046–1070.

Feng, Q., Z. Chen, C. Cheng, and H. Chang (2023, March). Impact of high- speed rail on high-skilled labor mobility in China. *Transport Policy 133*, 64–74.

Finnish Transport Infrastructure Agency (2024). Open api. https://vayla. fi/en/transport-network/data/open-data/api. Accessed: 2024-05-02.

Fintraffic (2024a). Nap - transport service catalog. https://finap.fi/#/. Accessed: 2024-05-02.

Fintraffic (2024b). Tms documentation. https://www.digitraffic.fi/en/ road-traffic/lam/. Accessed: 2024-05-02.

Fretz, S., R. Parchet, and F. Robert-Nicoud (2022). Highways, market access and spatial sorting. *The Economic Journal 132* (643), 1011–1036.

Fujita, M. and J.-F. Thisse (2002). Agglomeration and market interaction. *Available at SSRN 315966* .

Garcia-López, M.-À., A. Holl, and E. Viladecans-Marsal (2015). Suburbanization and highways in spain when the romans and the bourbons still shape its cities. *Journal of Urban Economics 85*, 52–67.

Gaubert, C. (2018). Firm sorting and agglomeration. *American Economic Review 108* (11), 3117–3153.

Gehlke, C. E. and K. Biehl (1934). Certain effects of grouping upon the size of the correlation coefficient in census tract material. *Journal of the American Statistical Association 29* (185A), 169–170.

Ghani, E., A. G. Goswami, and W. R. Kerr (2015, 04). Highway to success: The impact of the golden quadrilateral project for the location and performance of indian manufacturing. *The Economic Journal 126* (591), 317–357.

Gibbons, S., S. Heblich, and E. W. Pinchbeck (2024, September). The spatial impacts of a massive rail disinvestment program: The Beeching Axe. *Journal of Urban Economics 143*, 103691.

Gibbons, S., T. Lyytikäinen, H. G. Overman, and R. Sanchis-Guarner (2019). New road infrastructure: the effects on firms. *Journal of Urban Economics 110*, 35–50.

Gibbons, S. and S. Machin (2006). Transport and labour market linkages: empirical evidence, implications for policy and scope for further uk research.

Graham, D. J. (2007a). Agglomeration, productivity and transport investment. *Journal of transport economics and policy (JTEP) 41* (3), 317–343.

Graham, D. J. (2007b). Variable returns to agglomeration and the effect of road traffic congestion. *Journal of Urban Economics 62* (1), 103–120.

Graham, D. J. and K. V. Dender (2011). Estimating the agglomeration benefits of transport investments: some tests for stability. *Transportation 38*, 409– 426.

Graham, D. J. and P. C. Melo (2011). Assessment of wider economic impacts of high-speed rail for great britain. *Transportation research record 2261* (1), 15–24.

GraphHopper (2024). Documentation. https://docs.graphhopper.com/. Accessed: 2024-04-30.

Haapamäki, T., O. Harjunen, A. Kauhanen, V. Kuivalainen, K. Riukula, N. Valmari, and T. Väänänen (2024). Kasautumisen vaikutus tuottavuuteen: Työpaikkojen välisen saavutettavuuden vaikutus toimipaikkoihin ja työntekijöihin. *ETLA Report* (144).

Hanley, D., J. Li, and M. Wu (2022, March). High-speed railways and collaborative innovation. *Regional Science and Urban Economics 93*, 103717.

Hellerstein, J. K., D. Neumark, and K. R. Troske (1999). Wages, Productivity, and Worker Characteristics: Evidence from Plant-Level Production Functions and Wage Equations. *Journal of Labor Economics 17*, 409–446.

Helsinki Region Infoshare (2024). Hri-palvelu - avoimen datan palvelu. https://hri.fi/fi/hri-palvelu/. Accessed: 2024-12-05.

Heres, D. R., D. Jack, and D. Salon (2014). Do public transport investments promote urban economic development? evidence from bus rapid transit in bogotá, colombia. *Transportation 41*, 57–74.

Holl, A. (2004). Manufacturing location and impacts of road transport infrastructure: empirical evidence from spain. *Regional science and urban economics 34* (3), 341–363.

Holl, A. (2016, May). Highways and productivity in manufacturing firms. *Journal of Urban Economics 93*, 131–151.

HSL (2024). Liikkumistutkimus. https://www.hsl.fi/hsl/tutkimukset/ liikkumistutkimus. Accessed: 2024-09-17.

Imbens, G. W. and J. M. Wooldridge (2009). Recent developments in the econometrics of program evaluation. *Journal of economic literature 47* (1), 5–86.

Jones, P. (1979). New approaches to understanding travel behaviour: the human activity approach. In D. Hensher and P. Stopher (Eds.), *Behavioral travel modelling*. Redwood Burn Ltd.

Knudsen, E. S., K. Hjorth, and N. Pilegaard (2022, April). Wages and accessibility – Evidence from Denmark. *Transportation Research Part A: Policy and Practice 158*, 44–61.

Koopmans, C. and N. Mouter (2020). Cost-benefit analysis. In *Advances in Transport Policy and Planning*, Volume 6, pp. 1–42. Elsevier.

Koster, H. R., T. Tabuchi, and J.-F. Thisse (2022). To be connected or not to be connected? the role of long-haul economies. *Journal of Economic Geography 22* (4), 711–753.

Lazear, E., K. L. Shaw, G. E. Hayes, and J. M. Jedras (2022, December). Productivity and wages: What was the productivity-wage link in the digital revolution of the past, and what might occur in the ai revolution. Working Paper 30734, National Bureau of Economic Research.

Le Barbanchon, T., R. Rathelot, and A. Roulet (2020, December). Gender Differences in Job Search: Trading off commute against wage. *The Quarterly Journal of Economics 136* (1), 381–426.

Lee, J. K. (2021). Transport infrastructure investment, accessibility change and firm productivity: Evidence from the seoul region. *Journal of Transport Geography 96*, 103182.

Manning, A. and B. Petrongolo (2017). How local are labor markets? evidence from a spatial job search model. *American Economic Review 107* (10), 2877– 2907.

Martín-Barroso, D., J. A. Núñez-Serrano, and F. J. Velázquez (2020, January). The effect of accessibility on productivity in spanish manufacturing firms. *JOURNAL OF REGIONAL SCIENCE 55*, 102600.

Metsäranta, H., K. Riukula, A. Kauhanen, and P. Fornaro (2019). Liikennejärjestelmän työmarkkinavaikutukset ja niiden arviointi. *Valtioneuvoston selvitys- ja tutkimustoiminnan julkaisusarja*.

Mohammadi, M. Y., E. Rahimi, A. Davatgari, M. Javadinasr, A. K. Mohammadian, M. W. Bhagat-Conway, D. Salon, S. Derrible, R. M. Pendyala, and S. Khoeini (2023, July). Examining the persistence of telecommuting after the COVID-19 pandemic. *Transportation Letters 15* (6), 608–621.

Næss, P. (2006). Cost-benefit analyses of transportation investments: neither critical nor realistic. *Journal of critical realism 5* (1), 32–60.

Nilsson, J.-E. (1991). Investment decisions in a public bureaucracy: A case study of swedish road planning practices. *Journal of Transport Economics and Policy*, 163–175.

Oosterhaven, J. and T. Knaap (2017). Spatial economic impacts of transport infrastructure investments. In *Transport projects, programmes and policies*, pp. 87–105. Routledge.

OpenTripPlanner (2024). Product overview. https://docs. opentripplanner.org/en/latest/Product-Overview/. Accessed: 2024-04-30.

Pastinen, V., J. Piipponen, A. Supponen, L. Vuorio, J. West, A. Rantala, and T. Elolähde (2020). Helsingin seudun työssäkäyntialueen liikenne- ennustejärjestelmän kysyntämallit 2020.

Pekkarinen, T. (2006). Toimenpiteiden vaikutusten arviointi empiirisessä mikrotaloustieteessä. *Kansantaloudellinen aikakauskirja 102* (3), 307–320.

Pogonyi, C. G., D. J. Graham, and J. M. Carbo (2021, September). Metros, agglomeration and displacement. Evidence from London. *Regional Science and Urban Economics 90*, 103681.

Redding, S. J. and D. M. Sturm (2008). The costs of remoteness: Evidence from german division and reunification. *American Economic Review 98* (5), 1766–1797.

Redding, S. J. and M. A. Turner (2015). Transportation Costs and the Spatial Organization of Economic Activity. In *Handbook of Regional and Urban Economics*, Volume 5, pp. 1339–1398. Elsevier.

Ronkainen, T. (2024). Helsinkiläisten liikkumistottumukset 2023. *Helsingin kaupungin julkaisuja* (2024:9).

Rosenbaum, P. R. and D. B. Rubin (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika 70* (1), 41–55.

Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association 100* (469), 322–331.

Sanchis-Guarner, R. (2012). Driving up wages: The effects of road construction in great britain. Technical report, Centre for Economic Performance, LSE.

Sartori, D., G. Catalano, M. Genco, C. Pancotti, E. Sirtori, S. Vignetti, C. Bo, et al. (2014). Guide to cost-benefit analysis of investment projects. economic appraisal tool for cohesion policy 2014-2020.

Suomen ympäristökeskus (2024a). Elinympäristön tietopalvelu liiteri. https://www.syke.fi/fi-FI/Palvelut/Elinympariston_tietopalvelu_Liiteri/Mika_Liiteri. Accessed: 2024-12-05.

Suomen ympäristökeskus (2024b). Elinympäristön tietopalvelu liiteri. https://www.syke.fi/fi-FI/Palvelut/Elinympariston_tietopalvelu_Liiteri/Kayttoehdot. Accessed: 2024-12-05.

Tenkanen, H. and T. Toivonen (2020). Longitudinal spatial dataset on travel times and distances by different travel modes in helsinki region. *Scientific data 7* (1), 77.

Traficom (2024a). National travel survey. https://www.traficom.fi/en/ hlt. Accessed: 2024-09-17.

Traficom (2024b). Valtakunnallinen liikennemallijärjestelmä. https://www.traficom.fi/fi/liikenne/liikennejarjestelma/ valtakunnallinen-liikennemallijarjestelma?toggle= Valtakunnallinen%20henkil%C3%B6liikennemalli. Accessed: 2024-04-30.

Tveter, E., J. J. Laird, and P. Aalen (2022). Spatial aggregation error and agglomeration benefits from transport improvements. *Transportation Research Part A: Policy and Practice 164*, 257–269.

Tyndall, J. (2017). Waiting for the r train: Public transportation and employment. *Urban studies 54* (2), 520–537.

Tyndall, J. (2021). The local labour market effects of light rail transit. *Journal of Urban Economics 124*, 103350.

Van Wee, B. (2012). How suitable is cba for the ex-ante evaluation of transport projects and policies? a discussion from the perspective of ethics. *Transport Policy 19* (1), 1–7.

Venables, A., J. J. Laird, and H. G. Overman (2014). Transport investment and economic performance: Implications for project appraisal.

Venables, A. J. (2007). Evaluating urban transport improvements: cost– benefit analysis in the presence of agglomeration and income taxation. *Journal of Transport Economics and Policy (JTEP) 41* (2), 173–188.

Welde, M. and E. Tveter (2022). The wider local impacts of new roads: A case study of 10 projects. *Transport policy 115*, 164–180.

Xie, F. and D. Levinson (2011). Evaluating the effects of the i-35w bridge collapse on road-users in the twin cities metropolitan region. *Transportation planning and technology 34* (7), 691–703.

Zhang, X., Y. Hu, and Y. Lin (2020, January). The influence of highway on local economy: Evidence from China's Yangtze River Delta region. *Journal of Transport Geography 82*, 102600.

Zhu, S., D. Levinson, H. X. Liu, and K. Harder (2010). The traffic and behavioral effects of the I-35W Mississippi River bridge collapse. *Transportation Research Part A: Policy and Practice 44* (10), 771–784.

Appendices

Appendix A Empirical setting for studying the labour market effects of transport projects

Studying how transport projects affect employment and productivity requires a well-designed research framework, allowing for the comparison of individuals affected by the project with similar individuals that are unaffected by the project.[9] We demonstrate the perfect setting and deviations from it through the Rubin causal model, which structures our presentation of the empirical challenges. We consider the Rubin causal model in the case of a binary treatment to keep the exposition simple since our intent is to provide a unified framework to discuss challenges for causal inference. However, we note that many of the changes due to new transport infrastructure are continuous, such as changes in accessibility or travel times. Reducing these changes to a binary treatment variable can bias the estimate of the average treatment effect when the effect is non-linear. Therefore, using techniques that identify continuous treatment effects can provide a less biased estimates. The identification of continuous treatment effects is an active area of research with many recent contributions such as Callaway et al. (2024) and D'Haultfœuille et al. (2023). The Rubin causal model consists of a potential outcome framework and an assignment mechanism (Rubin, 2005). For ease of exposition, we consider the case of inferring causal outcomes for binary treatment $D \in 0, 1$. The potential outcome framework consists of units of observation i, covariates X_i, and potential outcomes $Y_i(D = 1)$ and $Y_i(D = 0)$, where the outcome $Y_i(D = 1)$ refers to the outcome of unit i when they are treated and the outcome $Y_i(D = 0)$ refers to the case when they are not treated. The causal effect β of the treatment for unit i is then defined from the two outcomes $Y_i(1)$ and $Y_i(0)$ usually as the difference $\beta = Y_i(1) - Y_i(0)$ but also from other definitions such as the ratio $\beta = Y_i(1)/Y_i(0)$.

To describe the causal effect of treatment, an assumption commonly referred to as the *stable unit treatment value assumption* (SUTVA). This assumption states that the outcome of unit i does not affect the outcomes of the other units $j \neq i$ (Rosenbaum and Rubin, 1983) and that the way the intervention under research is made does not affect the outcomes $Y_i(1)$ and $Y_i(0)$.

The other part of the Rubin causal model is the assignment mechanism, which controls the assignment of the treatment to the units (Rubin, 2005). Imbens and Wooldridge (2009) classify these assignment mechanisms into three categories: i) random experiments, meaning mechanisms that assign the treatment at random, ii) selection on observables, meaning mechanisms where the treatment probability does not depend on the potential outcomes, given the observable covariates of the units (the conditional independence assumption, Cox 1958), and iii) selection on unobservables, meaning all other mechanisms where the assignment probability may depend on the unobserved qualities of the units or potential outcomes of the treatment. The difficulty of analysing causal effects under each category of assignment mechanisms becomes progressively more difficult. However, for the first two cases, established and robust methods can be used to uncover the causal effects of treatment.

An optimal setup for studying the effects of transport investments involves randomising areas that receive transport projects. This would ensure that the areas receiving transport projects and those not receiving them would be similar in observable and unobservable qualities. Ideally, for SUTVA to hold, the research setting would also ensure that migration, relocation of establishments, or changes in traffic flows do not happen between the study areas or units. These kinds of conditions are impossibly restrictive, and in reality, researchers need to be mindful of potential violations of SUTVA and treatments that are assigned to observable or unobservable qualities of the units or outcomes.

[9] *This appendix chapter is partly based on Metsäranta et al. (2019).*

Appendix B Data sources in Finland

Transport networks. The Finnish Transport Infrastructure Agency (Väylävirasto) provides open data on Finland's transport network. They provide a map service and several APIs through which it is possible to access a plethora of information about road, rail, water, and air transport networks (Finnish Transport Infrastructure Agency, 2024). The oldest snapshot of the road network was obtained in 2019. The Finnish road network also hosts traffic measurement points that monitor traffic volume and speed at over 450 locations throughout the road network. These data are provided in real time, as well as historical records of the measurements. It is also possible to access the raw data from these measuring points, which include every measurement that has been made (Fintraffic, 2024b).

Data on the public transport network and timetables are provided in the General Transit Feed Specification (GTFS) format. Finnish GTFS data are scattered across providers but are usually provided by city authorities, as in Tampere, or a city's public transport authority, such as HSL in the Helsinki region or Föli in the Turku region. National Access Point managed by Fintraffic gathers the sources of GTFS information from all public transport providers to a searchable database (Fintraffic, 2024a). Historical GTFS data are also available. For example, the GTFS feeds of the Helsinki region can be found from 2016 onwards.

Information about historical roads and rail networks is typically gathered from regional historical maps. National Land Survey of Finland has gathered resources of old maps of Finland to their website[10]. These maps are mostly in picture form and thus require technical solutions to be made useful for empirical work. Information about the history of Finnish road network is scattered, but an individual has gathered this history to his private website[11]. The information is largely unsourced, but can serve as a starting point for finding historic roads or old alignments of current roads.

Travel times. In addition to travel demand models that assess the number of trips, modes, destinations, and routes, there are tools to calculate travel times for trips in a certain mode. Tools such as OpenTripPlanner (OpenTrip- Planner, 2024) and GraphHopper (GraphHopper, 2024) are routing tools that use open-source transport network data from OpenStreetMap to calculate the fastest routes and travel times between given coordinates. These tools use commonly available GTFS data to provide public transport routes and travel times. Digitransit is a routing service that collects different GTFS feeds from Finland to provide a routing service for public transport (Digitransit, 2024). They also provide an API to create these routing requests programmatically. Google Maps and Bing Maps provide routing services that can be used to calculate routes and travel times between coordinates. These routing tools are useful for calculating travel times between areas, and the Google Maps API can even consider congestion based on historical traffic information. These tools can also be used in other countries.

An additional resource for travel times in the Helsinki region is the Helsinki Region Travel Time Matrix, which was developed by a research group at Helsinki University (Tenkanen and Toivonen, 2020). This travel time matrix contains the travel times by walking, cycling, public transport, and car for the years 2013, 2015, and 2018 between all Statistics Finland 250x250m grid cells.

Travel demand models. Some areas, such as the Helsinki, Tampere, and Turku regions, have travel demand models that can be used to model the effects of transport projects on travel times. The Helsinki region travel demand model is the only model that has been extensively documented (Pastinen et al., 2020), and information on other models is scattered across different documents. The use of these models requires a licence for proprietary software, as route choice models are typically handled by third-party software such as EMME or VISUM.

[10] *https://www.maanmittauslaitos.fi/suomenvanhatkartat*

[11] *https://www.mattigronroos.fi/w/index.php/Etusivu*

The Finnish Transport and Communications Agency is developing a national travel demand model system that can be used to assess transport-related effects of transport projects across the entire country (Traficom, 2024b). The model system is supposed to make use of local travel demand models and form a comprehensive system of different travel demand models for intra-regional and inter-regional transport forecasting and analysis.

Travel surveys. Travel surveys are an excellent source of information on individual travel behaviours. Three main travel surveys are conducted in Finland. The national travel survey by Traficom (Traficom, 2024a), the Helsinki region travel survey by HSL (HSL, 2024) and the Helsinki city travel by the City of Helsinki (for example, Ronkainen, 2024). As the names suggest, the national travel survey surveys the entire nation, and the Helsinki region travel survey surveys citizens in the Helsinki region. These two surveys are conducted approximately every four years. The Helsinki travel survey, however, only surveys people living in Helsinki and is conducted annually.

Other related data sources. Other sources of spatial data related to transport and land use are the Liiteri and Helsinki Region Infoshare (HRI). Liiteri is a service by the Finnish Environmental Institute that gathers spatial data on the built environment and zoning Suomen ympäristökeskus (2024a). This service includes many datasets, but access to some of the data requires a paid licence Suomen ympäristökeskus (2024b). HRI is a similar service for data on municipalities in the Helsinki Region Helsinki Region Infoshare (2024). HRI includes, for example, data on the average traffic in municipalities, parking places in municipalities, and data about the population living in the Helsinki region.