UNIVERSITA' DEGLI STUDI DI PADOVA

DIPARTIMENTO DI SCIENZE ECONOMICHE ED AZIENDALI
"M. FANNO"

CORSO DI LAUREA IN ECONOMIA

PROVA FINALE

"THE EFFECTS OF TRANSPORT INFRASTRUCTURE ON ECONOMIC DEVELOPMENT: A RECENT LITERATURE REVIEW"

RELATORE:

CH.MO PROF. ROBERTO BONFATTI

LAUREANDO: GIOVANNI SCAPOLO

MATRICOLA N. 2003658

ANNO ACCADEMICO 2022 – 2023

Dichiaro di aver preso visione del "Regolamento antiplagio" approvato dal Consiglio del Dipartimento di Scienze Economiche e Aziendali e, consapevole delle conseguenze derivanti da dichiarazioni mendaci, dichiaro che il presente lavoro non è già stato sottoposto, in tutto o in parte, per il conseguimento di un titolo accademico in altre Università italiane o straniere. Dichiaro inoltre che tutte le fonti utilizzate per la realizzazione del presente lavoro, inclusi i materiali digitali, sono state correttamente citate nel corpo del testo e nella sezione 'Riferimenti bibliografici'.

I hereby declare that I have read and understood the "Anti-plagiarism rules and regulations" approved by the Council of the Department of Economics and Management and I am aware of the consequences of making false statements. I declare that this piece of work has not been previously submitted – either fully or partially – for fulfilling the requirements of an academic degree, whether in Italy or abroad. Furthermore, I declare that the references used for this work – including the digital materials – have been appropriately cited and acknowledged in the text and in the section 'References'

Giovanni Scapolo

Abstract English

In the last decade, there has been growing interest in the causal effects of transport infrastructure on various aspects of economic development. This has been possible thanks to innovations in theoretical designs, and data availability. In this work, after analyzing what the before-mentioned innovations consist of, I try to regroup the most important works that shaped the research field, as well as the biggest fallacies that researchers had to and will face in gaining a deeper understanding of the causal relationship between transport infrastructure and economic development.

Abstract Italiano

Nell'ultimo decennio si è verificato un cresciuto interesse per gli effetti causali delle infrastrutture di trasporto su diversi aspetti di sviluppo economico. Ciò è stato possibile grazie ad innovazioni in modellazioni teoriche e alla disponibilità di dati. Dopo aver analizzato quali sono queste innovazioni, in questo lavoro, cerco di raggruppare i lavori più importanti che hanno plasmato il campo della ricerca, così come le principali fallacie che i ricercatori hanno dovuto e dovranno affrontare per ottenere una comprensione più profonda della relazione causale tra infrastrutture di trasporto e sviluppo economico.

Table of Contents

Words: 9985

1. Introduction

In a world where a billion people live without access to any type of reliable infrastructure and many more have access to poorly maintained ones (Wenz et al, 2020), policymakers need to address how investment in infrastructure is to be allocated. For example, one question that developing countries have to ask themselves is whether it is more beneficial to improve connectivity in already connected areas or to link more communities with lower-quality connections.

The debate around efficiency in allocation in this field has persisted for more than thirty years. Since Aschauer's first seminal paper in 1989 (Aschauer, 1989) on the question of the productivity of public expenditure on transport infrastructure, the literature has made great strides, arriving at an almost unanimous consensus that infrastructure is a key element for the development of any economy (Foster et al, 2023a). Nonetheless, in the following decades, researchers had yet to answer crucial questions regarding policy-makers such as the magnitude of development directly attributable to these investments, or the specific effects of infrastructure on the labour market or firm competitiveness, to name just two.

These gaps in the research have been caused by the impossibility of gathering and elaborating enough micro-economic data to conduct studies that obtained credible causality in their results. However, in the last decade, these gaps have gradually started to be filled with the conjoint advent of new technologies for data usability and accessibility, such as geospatial data first introduced in the field by Anas and Liu (2007), and the advancements in the econometric instruments at the researchers' disposal (Redding and Rossi-Hansberg, 2017). These developments have spawned a second wave of literature that is based on more granular data and, for this reason, is more bespoke in the methodologies and targeted in its findings. This newfound ability has given the researchers the chance to deepen their understanding of entire sub-categories that focus on specific developmental outcomes of infrastructure investments.

The objective of this work is to provide a generalized look at the results obtained in the transport infrastructure research field during the last decade, with a focus on research that managed to resolve econometric challenges, innovated in data usage techniques, or studied new-found developmental outcomes of transport infrastructure. While a predefined geographical scope was not set, the literature is mostly focused on developing countries where gaps in connectivity are the largest, and infrastructure has the potential to generate the most significant outcomes.

The remainder of the paper is structured as follows. Section 2 will define the search and selection criteria for the literature review. In section 3 I will discuss the issue of endogeneity and which methods researchers have used to tackle it. Section 4 will be devoted to the analysis of several foundational papers that have brought innovations to the field. In section 5 I will discuss the issue of publication bias; while in section 6 there will be a brief conclusion.

2. Search & Selection Criteria

The following review has been conducted by first gathering a vast amount of works within the literature, then applying a series of below-described filters to obtain a database of key papers that have been particularly significant for their approach or their result.

To do this, the selected time frame has been set from 2010 to 2023, given the aforementioned increase in the microeconomic developmental outcomes studied. The geographical scope has been dictated by the literature: while a small number of seminal papers are focused on developed economies, the majority study low and middle-income countries. Furthermore, papers that were not solely addressing transport-related infrastructure have also been scrutinized.

In effect, the most stringent filter has been the econometric method(s) used by researchers to tackle the issue of endogeneity in the allocation of infrastructure and how credible it was at resolving it. And while this issue will be discussed later in the paper, the analysis did not exclude non-peer-reviewed articles *ex-ante*, however, the publication has shown to be a reliable proxy for credible approaches. Nonetheless, it must be noted that studies that explicitly address endogeneity do not appear to yield significantly different results from those that will be analyzed (Foster et al. 2023b).

The survey of analyzable papers is composed of the conjunction of previously made systematic literature reviews and meta-analysis papers on the topic that had similar selection characteristics with the before-described ones. The foundation from which the other works have been scrutinized was composed of the two joint papers that V. Foster et al. published in 2023: one as a systematic review (Foster et al, 2023a), and the other as a meta-analysis paper (Foster et al, 2023b). From these two works, all related secondary literature, which extended to more than 300 research papers, was examined, and those selected reviews are listed below in Table 1.

Table 1: Scrutinized review papers

Author(s) (Date)	Relationship studied	Nr. of studies analyzed	Selected Time-Frame
Redding and Turner (2015)	Spatial distribution of economic activity and transport costs	105	1958-2014
Redding and Rossi-Hansberg (2017)	Quantitative Spatial Economics	126	1967-2016
Berg et al. (2017)	Transport Policies and Development	138	1993-2017
Deng (2013)	Transport Infrastructure and Productivity and Economic Growth	48	1989-2012

3. Endogeneity in Allocation

3.1 The Issue of Endogeneity

One key issue that repelled the development of the research field until the 2010s was the inability to obtain results that were credibly unbiased from the problem of endogeneity in the allocation of transport infrastructure. These biased results halted the finding of a credible causal relationship between infrastructure and economic outcomes studied, inhibiting the possibility of drafting policies that efficiently allocate one of the biggest sources of government expenditure, especially for developing economies (ITF, 2021).

The selection bias concerns derive from different sources but all of these have as their originating moment the drafting of infrastructure projects. This spurs from the fact that the planners intend to boost the endowment of infrastructure in general, but due to various constraints, they inevitably have to prioritize certain areas. This process of prioritization generates the so-called "non-random allocation patterns". One such constraint could be economic: where there is higher economic potential, infrastructure investments may be prioritized. At the same, opposite political constraints could arise: areas that policy-makers see as preferable for political reasons might receive investments disproportionally. Furthermore, the choice could be altered by lobbying efforts and/or budgetary constraints that determine a non-efficient allocation of these funds.

Alas, endogeneity could also originate from the phenomenon of reverse causality: this occurs when an already-developed area attracts more than proportional investments compared to its economic state. The understanding of this circumstance is particularly intriguing to study because of its policy outcomes. Comprehending whether the endowment of infrastructure boosts economic development or, on the contrary, the prospects of development are the determinants of infrastructure investments is paramount to the whole research field.

3.2 Endogeneity-Tackling Methods

As previously discussed, the biggest challenge is that regressions that try to identify the distinct causal relationship between improvements in the endowment of infrastructure and various micro- and macroeconomic indicators (employment, land value, household income, firm competitivity,…) have to account for selection bias. In particular, Ordinary Least Squares (OLS) regressions cannot consistently estimate the causal effects because of the non-random selection of transport routes.

The main approach used by researchers is the conjoint development of instruments and control variables that credibly manage to identify the causal effect directly caused by transportation improvements. And while program evaluation literature indicates other auxiliary approaches, which almost always remain in the domain of quasi-experimental designs, these have been less applied in this empirical literature. Nonetheless, I will still briefly discuss them later in section 3.3.

To better elaborate a general regression that includes the aforementioned instrumental variable approach, we exploit Redding and Turner's (2015) theoretical formalization of intercity regression. Jot that, while encompassing other theoretical differences, the same analytical form can be achieved for intracity regression. Nonetheless, first-stage regressions for intercity regression usually follow the specification form:

$$E_{it} = D_0 + D_1 d_{it} + D_2 z_{it} + \eta_i + y_t + u_{it} \tag{3.1}$$

Here E_{it} is the economic outcome of interest, d_{it} are the time- and location-variant control from equation (3.2); z_{it} are the instruments or excluded exogenous variables; η_i are the location-specific, time-invariant unobservables; y_t are the time indicators; and u_{it} is a time-varying location-specific residual.

Meanwhile, formal second-stage regressions that the researchers postulate are similar to the following:

$$L_{it} = C_0 + C_1 d_{it} + C_2 x_{it} + \delta_i + \theta_t + \epsilon_{it} \tag{3.2}$$

Let L_{it} be defined as the economic outcome of interest for location *i* at a specific time *t*, such as employment level, population growth, land value, or household welfare. x_{it} is a vector of location and time-specific covariates, and finally, d_{it} denotes the transportation variables that are being analysed, such as road, rail, or public transport mileage. Furthermore, δ_i denotes location-specific time-invariant unobservables, θ_t indicates a time effect for all locations, inside and outside the analysed network. Lastly, ϵ_{it} denotes the time-varying but location-specific residual. The coefficient of interest is C_1, which measures the effect of inter-city transport costs between city *i* and another unit.

When the two-stage equations (3.1 and 3.2) are combined, it is possible estimating the causal effect of infrastructure on the desired economic outcomes through the coefficient C_1. To do this, we can use the most common method of the "two-stage least squares" (TSLS). The main requirements to use the TSLS method are that the instruments significantly differ from zero in the first-stage regression ($D_2 \neq 0$) and that the same instruments interfere with the analyzed economic outcomes only though the controls, meaning that $cov(\epsilon_{it}, u_{it}) = 0$

Whilst using the instrumental variable approach, literature has primarily developed around three main instrumental variables strategies. In section 4 at least one paper per approach will be analysed in order of publishment date. The first, the planned route I.V. approach, is a strategy reliant on planning maps and projects as a source of quasi-random variation in the observed infrastructure. The second, the historical route I.V. approach, uses transportation corridors that predate the analysed period as a source of quasi-random variation in the observed infrastructure. The third, the inconsequential place approach, is based on selecting a sample of infrastructure that is inconsequential. This means that the unobservable attributes that create the non-random allocation patterns do not impair the placement of infrastructure.

The implementation of these strategies heavily depends on how they are implemented in the specific case, and researchers have not reached a consensus on which is the most efficient method, or which are the prerequisites to use one when the setting permits the use of multiple approaches.

3.3 Complementary Approaches

As previously stated, program evaluation literature has suggested the use of other methods of analysis that have not gained major traction in the literary corpus. However, it is not uncommon for researchers to use designs that combine multiple methods during the same exercise to achieve the most robust result possible.

Regression Discontinuity Design (RDD) can be used when infrastructure allocation or policy has a specific threshold or cut-off point that can be identified. By examining the possible presence of discontinuities in outcomes around the threshold, this method identifies the causal effect of infrastructure on economic outcomes. This approach exploits the artificial variation created by the threshold to establish causality in some particular settings.

The Matching Estimators technique aims to create comparable groups by matching units based on observable characteristics. This is particularly useful when the treatment and control groups are heterogeneous but, units within the groups have other observable similarities. By matching these units, there is a reduction in differences between the treated and control groups, helping to address endogeneity among other problems.

Another used technique is based on fixed effects and panel data models, which are both used to control for unobserved heterogeneity and time-invariant factors that may influence infrastructure allocation and economic outcomes. By including fixed effects, these models can account for example time-constant factors that could introduce bias.

Moving on to the Difference-in-Difference(s) (DiD) analyses are used to compare changes in outcomes between treated and control groups before and after infrastructure investments took place. The concept behind these analyses is the assumption that the two units would have had similar trends in the absence of infrastructure. In this way, DiD finds a way to isolate the causal effect of infrastructure on the studied economic outcomes.

At last, Structural Equation Modelling (SEM) is an approach which incorporates both measurement and structural models to examine the relationships between infrastructure allocation, latent variables, and economic outcomes. To do this, a model which encompasses the three before-mentioned factors has to be developed by identifying, and estimating the latent variables. When correctly specified, SE Models allow researchers to estimate direct and indirect effects and control for measurement errors and endogeneity.

4. Analysis

In the following section, I will analyse some of the most important papers in the field that contributed to the discipline in a particular fashion. These contributions have several provenances: they could have introduced an innovative method of addressing endogeneity concerns, they managed to tie in a single model with previously unconnected dimensions, or they obtained significant policy-relevant results.

4.1 Donaldson: Railroads of the Raj

In its seminal paper "*Railroads of the Raj: Estimating the Impact of Transportation Infrastructure*" (2018), Donaldson analyses the improvements that the 67-thousand kilometre railway network built in Colonial India between 1853 and 1930 brought to both trade-openness and district-level real incomes, leading to overall welfare gains. To do this, he sourced data from the historic archives of the British Empire of all infrastructure projects in the *Raj*, which provided unprecedented depth for some recorded data, such as inter-district trade flows.

To correctly specify such outcomes he expanded an already established general equilibrium model theorized by Eaton and Kortum (2002)[1]. Having said that, the analysis was structured in four steps with each obtaining intermediary results to finally achieve the causal estimation of welfare gains caused by an increase in trade. The first two steps were dedicated to estimating the magnitude of infrastructure construction on the trading openness calculating how it lessened trade costs and boosted interregional trade flows. Secondly, he assessed the direct welfare gains, calculated in real income levels, from infrastructure construction. Lastly, he connected the previous steps to calculate what percentage of welfare gain generated by the railroads was to be attributed to the decrease in trade costs.

---

[1] Which was, in turn, theoretically founded on the Ricardian trade model

Focusing on the findings of the second step, which is the one of interest, the theoretical model used is based on the assumption that welfare is equal to real income and, that is comparable to income per unit of land area which, in turn, is given by real land rents. These are calculated by diving average land rents (r) by a weighted price index of commodities (P).

Moving to the empirical analysis, Donaldson constructs an OLS regression (this choice will be discussed later), that connects the theory by which welfare is equal to the real agricultural income per acre in a specific district o in a given year t to the connection of the district to the railway network. The regression takes the following form:

$$\ln\left(\frac{r_{ot}}{P_{ot}}\right) = \beta_0 + \beta_1 + \lambda rail_{ot} + \varepsilon_{ot} \tag{4.1}$$

Furthermore, in this equation, the regressor $rail_{ot}$ is a dummy variable which gets activated when the railway connection happened and, β_0 are the district fixed effects, while the variable of interest is β_1.

Coming back to the choice of utilizing OLS regression, the author provides a solution to the issue of non-random allocation patterns. He avoids using an instrumental variable approach because, in the eyes of the writer, the problem could be discounted altogether due to the exclusively military-driven allocation patterns. Nonetheless, he addresses this hypothetical issue with a DiD analysis that is founded by comparing the "effects" of railroads where an additional 40 thousand kilometres of railroads were planned but not built, compared to districts that were completely unaffected by railroad plans. And, by finding that the two groups of districts did not significantly develop in different ways, Donaldson demonstrates that selection bias did not influence the estimations of his study. It is important to note that, while this endogeneity-resolving strategy does not rely on instruments, its theoretical foundation derives from Baum-Snow's (2007) "planned route instrumental variable approach", of which Donaldson's paper is the most influential exponent.

The results found from equation (4.1) by Donaldson highlight that the connection to the railway network has caused a growth in real income by 16.4%. Furthermore, in the rest of his analysis, he also finds that this growth can be attributed to the falling trade costs of 52%.

4.2 Duranton & Turner: Urban Growth and Transportation

Duranton and Turner (2012) in their paper: "*Urban Growth and Transportation*" design an instrumental variable approach to tackle the issue of endogeneity in the allocation of infrastructure investment. They utilize this method to estimate some of the parameters present in their theoretical model based on simultaneous equations that managed to link the growth of the endowment of infrastructure, especially highways, with the growth of employment levels in a city. To conduct this study they researched the 227 US metropolitan statistical areas (MSAs) during the periods 1963-1983 and 1983-2003.

Looking at the theoretical framework of simultaneous equations, the authors managed to build one on only three equations that presented nine structural parameters: α (share of transportation costs in household expenditure), β (share of land in expenditure), δ (elasticity of unit transportation costs to road provision), λ (employment adjustment factor), η (elasticity of current roads to past employment), θ (road adjustment factor), σ (agglomeration economies), φ (land supply elasticity), and w (rural wage).

Coming back to the method used to resolve selection bias, the authors recognize that it might be originated from a combination of factors such as: reverse causality, omission of variables related to the differences in the initial stock of infrastructure, or, more simply, observation errors. For this reason, the authors adopt three instruments: the kilometres of Interstate Highways planned during the project of 1947, the kilometres of railways active in 1898, and the routes of major expeditions of exploration between 1528 and 1850. These instruments fall under the "historical route instrumental variable approach" that relies on much older routes than those analysed and treats them as a source of quasi-random variation observed in observed infrastructure (Redding and Turner, 2015). This approach first pioneered in this paper, has been replicated in its methods by many other illustrious researchers such as Baum-Snow et al. (2017) and Garcia-Lopez, Holl, and Viladencas-Marsal (2015).

Together with these instruments, the authors include in their regressions a series of explicative variables to enhance the robustness such as the levels of growth antecedent to the analyzed periods, geographical effects, socioeconomic divisions, and census divisions.

Interestingly, the regressions are treated using the less common method of "Limited Information Maximum Likelihood" (LIML) as it permits to obtain more reliable estimates when in the presence of weak instruments. Nonetheless, to be used this method requires some unique assumptions such as the presence of valid instruments and that the model is correctly specified[2].

To obtain the results, the researchers assigned the values to each of the nine parameters either from estimates obtained through regressions or from estimates from the literature. Furthermore, to assure the validity of the results, the authors carried out a check using the "Generalized Method of Moments" (GMM) treatment. This has obtained non-significantly different results from the estimates achieved with the LIML method.

The results reached by Duranton and Turner revealed that a 10% growth in the stock of transport infrastructure has increased employment levels by 15% during a twenty-year study period. At the same time, it is pointed out that a city that has its infrastructure stock grown by 10% in the period will receive less infrastructure growth in the following twenty years by 27%.

4.3 Faber: Trade Integration, Market Size, and Industrialization

In Faber's (2014) paper "*Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System*" the objective of the analysis was to estimate the direct economic effects on China's rural counties that were randomly affected by the construction of the NHTS network[3]. In particular, his study focuses on how the connection to major production centres caused shifts in population, occupation, government revenue, and industrial and agricultural production. Construction was bolstered by the provincial government's investments that reduced the end of the project to 2007 with three intermediary periods (1992-1997,1997-2003, and 2003-2006) that provide a deeper understanding of the incremental nature of the changes caused by the connection to the infrastructure.

In contrast with the first two analyzed studies, in his search for a causal correlation between the introduction of infrastructure and the aforementioned economic outcomes, Faber's first endeavour is the resolution of selection bias between the targeted metropolitan areas. To do this, he constructs two hypothetical alternative route placements that have as founding assumption exogeneity in allocation. The first is a combination of two algorithms: a least-cost route placement algorithm built with GIS data, and a minimum-spanning tree network. The second is a Euclidean network tree of straight-line bilateral connections. While yielding different results,

---

[2] To verify this the authors perform a Hansen J test which verifies the conjoint exogeneity of the instruments to the first stage regressor.

[3] The network was built between 1992 and 2020 to connect all cities with more than 500 thousand inhabitants

both of them are instruments of interest which were used conjointly since the least-cost path network yielded more precise route placements while the Euclidean one was more accurate in the effectively built bilateral connections.

These two networks were used as instruments in a strategy defined as the "*inconsequential place approach*" which has as its founding assumption that if it is possible to identify geographic areas that are economically insignificant to the targeted cities, in those areas planners will not be incentivized to steer away from the least-cost route placement. In Faber's study, this assumption was verified due to the average market size of the studied locations being 1/25th compared to those of targeted cities. While standing as a pivotal paper on how market access impacts development, this paper utilizes an instrumental variables approach that was first introduced by Chandra and Thompson (2000).

The generalized first-stage regression takes also into account a series of province-level fixed effects, such as the shares of tertiary education and urban population before the effects caused by the highway. These further reinforce the robustness of the study findings, and in the taxonomy is compacted into the vector X_{ip}. The regression takes the following form:

$$\ln\left(y_{ip}^{2006}\right) - \ln\left(y_{ip}^{1997}\right) = \lambda_p + \beta Connect_{ip} + \eta X_{ip} + \epsilon_{ip} \qquad (4.2)$$

where $Connect_{ip}$ is the variable of interest for the connection to the highway in the county *i* of province p, while λ_p encompasses the province-level fixed effects.

Following this econometric analysis, Faber finds those counties that were incidentally connected to the network suffered a reduced economic growth compared to the non-connected ones by 19% which was mainly attributable to a reduction in industrial output of 26%, meanwhile effects on population reallocation were non-significant.

After assessing the magnitude of the effects of the infrastructure network, the author starts to review the economic forces that might be in action in this setting to determine which best explains these results. The comparison revolves around two competing mechanisms: one driven by inter-regional reductions in trade costs between the studied counties and metropolitan areas, and another based on the decentralization of economic activities from connected to non-connected peripheral counties. To do this, a series of additional estimations are conducted and the results are in line with the first mechanism which also explains why connected counties closer to targeted cities suffer significantly more than more remote connected counties. This result aligns with numerous economic theories that could explain the phenomenon such as increasing returns to scale, monopolistic competition, self-reinforcing agglomeration forces, comparative disadvantages, and many more.

4.4 Ghani, Goswami, and Kerr: Highway to Success

Another paper that focuses on the economic development brought by the introduction of highways is Ghani, Goswami, and Kerr's (2016) "*Highway to Success: The Impact of the Golden Quadrilateral Project for the Location and Performance of Indian Manufacturing*". This study differs from Faber's since it focuses on a set of microeconomic outcomes related to the performance of the organized manufacturing industry in India such as labour productivity, market access, and product range. This specific analysis could be achieved only due to the granular data coming from the first two out phases of the Golden Quadrilateral Project (GQ). The datasets came from different stages of the project: before the plan was put into place in 1994, to the years were works began in 1999, to the aftermath of the project in 2009.

To confront the project's effects, the authors created two categories of businesses: those that are under 10 kilometres away from the infrastructure, and those that are 10 to 50 kilometres away from the highway. Meanwhile, those firms that had their location in the four nodal cities, and those that were more than fifty kilometres away were kept only as control groups.

One interesting econometric approach that this study takes is that of long-differenced estimates: a panel-data analysis technique used to address the issues of endogeneity caused by time-invariant factors. By comparing all the variable's values in two, significant moments in the history of the project, the authors managed to do just that. In this case, the comparison was made between the 1999 and 2009 datasets, nevertheless, the 1994 to 2009 comparison obtained akin results, and the general specification took the following form:

$$\Delta Y_i = \sum_{d \epsilon D} \beta_d \, X(0,1) GQDist_{i,d} + \gamma X_i + \varepsilon_i \qquad (4.3)$$

where d is the set of districts that are in one of the two aforementioned groups based on distance D, X_i is a vector of district-level time-variant controls, while β_d is the variable of interest.

The analysis of the results proceeds with different steps that add an increasing amount of controls being accounted for. Nevertheless, the findings remain relatively stable during the process and always show an overall improvement for the whole sector. These improvements are significantly more accentuated for the dataset of newly-formed firms. Other interesting findings are that the level of employment remained unchanged by the project and that districts that are in the control groups, meaning more than 50 kilometres away or reside in the nodal cities, do not suffer from the opening of the highways.

Another singular method to check the robustness of the aforementioned findings was provided by the setting. In fact, during a similar period, another mostly non-incidental highway project was planned: the NorthSouth-EastWest Project (NS-EW). However, for multiple reasons, the project was built with more than a decade of delay compared to the GQP, but the policy setting for the allocation of the two infrastructure projects remained the same. This allowed researchers to directly compare, while still attached to some time-variant controls, the two district groups and the results tell a story of lower estimates with much higher standard errors, obtaining almost always non-significant results.

While the comparison with the NS-EW Project might seem reassuring regarding the robustness of the findings, the authors still proceed with another check based on the fears that the dataset of the competing project might have been too reduced or the policy setting may have differed from the GQ Project. For this reason, the researchers proceed with a straight-line IV approach, similar to the Euclidean setting before described by Faber. The findings with this approach are non-significantly different for most economic outcomes to the long-differenced method, with an elasticity of 0.43 for the overall output of the manufacturing sector.

After assessing the effects of the GQ Project by estimating various microeconomic indicators, the authors proceed with dynamic specifications of said outcomes yearly, finding that the effects start to be noteworthy when 80% of the network was built.

4.5 Asher & Novosad: Rural Roads and Local Economic Development

Asher and Novosad (2020) in their paper "*Rural Roads and Local Economic Development*" focus on finding the effects on several economic outcomes such as employment patterns and economic output that the construction of paved rural roads had on remote villages. To accomplish this they restricted their search to Indian villages that were connected with paved roads during the Prime Minister's Village Road Program (PMGSY), one of the biggest endeavours of rural road construction in the last century which aimed at linking, from 2000 to 2015, more than 185 thousand villages. The data came from tens of matched datasets regarding the project's household-level and firm-level rural economic situations.

Importantly for the construction of the paper, the program guidelines dictated that bigger villages had to be prioritized, and the defining thresholds were set from the population census.

In fact, excluding particular areas, villages with more than a thousand inhabitants were connected in the first phase; and during the second phase, all villages with more than 500 inhabitants were connected. Even though road placement may have been impacted on a case-by-case scenario by political or economic factors, these did not impact the population thresholds imposed by program guidelines. This meant that, throughout the program, the probability of receiving a road was discontinuous around the thresholds, permitting the use of a fuzzy regression discontinuity design to conduct this study.

Having assessed the choice of design, the econometric analysis was based on an IV specification that used population bands as instruments[4]. To use these as instruments the authors had to satisfy certain requirements regarding the distribution of the variable such as the absence of discontinuity. Once verified, the following form of first-stage (4.4) and second-stage regression (4.5) is then used:

$$Y_{vj} = \beta_0 + \beta_1 Road_{vj} + \beta_2 (pop_{vj} - T) + \beta_3 (pop_{vj} - T)\ X\ 1\{pop_{vj} \geq T\} + \xi X_{vj} + \eta_j + \upsilon_{vj} \qquad (4.4)$$

$$Road_{vj} = \gamma_0 + \gamma_1 1\{pop_{vj} \geq T\} + \gamma_2 (pop_{vj} - T) + \gamma_3 (pop_{vj} - T)\ X\ 1\{pop_{vj} \geq T\} + \nu X_{vj} + \mu_j + \upsilon_{vj} \qquad (4.5)$$

In these equations Y_{vj} is the ultimate variable of interest in a village v which resides in a population bandwidth group j, T is the threshold, pop_{vj} is the baseline village population, X_{vj} measures a set of village controls, while μ_j and η_j are district-level threshold fixed effects. Village-level controls account for the baseline presence of public amenities, the land area dedicated to agriculture, the literacy rate, and much more. The fuzzy nature of the design means that, in contrast to a defined bandwidth of studied villages, the paper implements a method that gives more weight to observations that are closer to the threshold.

The economic outcomes of interest were estimated through five indexes: transportation services; allocation of labour; employment levels in non-agricultural village firms; agricultural investments and yield; income and assets. The results undoubtedly estimated positive effects on the transportation services of 48%, meanwhile, there is a notable reduction in employment in the agricultural sector (34%), with a slightly lower growth in employment in non-agricultural

[4] as indicated by Imbens and Lemieux (2008)

firms (27%). Together with the non-significant increase in agricultural output, it is possible to paint a setting where the roads meant an intersectoral reallocation of labour, without meaningful consequences on agricultural production. In turn, this could be achievable with an increase in technological inputs. Lastly, the index regarding consumption and assets does not show significant effects.

To corroborate the findings, the authors proceed with a series of robustness checks such as a placebo exercise which used a database of villages within states that did not follow the program guidelines. All these checks show that: there are no non-specified variables, the threshold does not withhold any intrinsic endogenous effect, and there were no causal changes in permanent migration patterns.

4.6 Morten & Oliveira: The Effects of Roads on Trade and Migration

Another more recent paper that studies the role of transport infrastructure in the displacement of labour is Morten and Oliveira's (2018) "*The Effects of Roads on Trade and Migration: Evidence from a Planned Capital City*". More specifically their paper has as its objective the understanding of what is the magnitude of welfare gains that occur when internal displacements happen after the construction of roads. This study was possible due to the particular setting which was provided by Brazil's initiative to build a new capital city more geographically centred during the 1960s, thus allowing for a new network of radial roads to be built with state capitals.

However, before comprehending the effects on welfare, the authors proceed to estimate migration patterns. To do this they utilize gross migration flow data, which permitted them to separate the migration phenomenon caused by shifts in wages and prices, from the one directly related to the roads. The empirical approach used to address endogeneity was the minimum spanning, straight-line IV approach first proposed by Chandra and Thompson (2000).

The approach was put into place by hypothesizing that the only policy goal was to connect Brasilia radially to the state capitals with only two highways per cardinal direction. Once formed, the travel times of the hypothesized network were confronted with the travel times as if the network did not exist. To do this, they specify the following equation:

$$y_{odt} = \gamma_{ot} + \gamma_{dt} + \gamma_{od} + \alpha_t RTT_{od} + \upsilon_{odt} \tag{4.6}$$

where RTT_{od} is the time-invariant log reduction in travel times in city pairs o to d obtained by introducing the hypothesized network; γ_{ot}, γ_{dt}, and γ_{od} are respectively origin-year, destination-year, and origin-destination fixed effects. Meanwhile, the dependent variable y_{odt} embodies the variables of interest: increase in trade or migration. This phase of the work highlights that the road network decreased migration costs by 8% and trade costs by 9%, with more significant results for cities farther from Brasilia.

After assessing the magnitude of migration flow changes after the construction of a road, the authors calculate the welfare gains that are attributable to the road. With welfare gain regarding migration, they consider all the reduction in costs that an immigrant pays compared to a comparable individual who did not migrate from his home region. To correctly specify these amounts Morten and Oliveira include not only the one-time migration costs, but also subsequent costs such as the ones of visits, and the ones relating to the "caloric tax"[5].

Lastly, they proceed to estimate the decomposed nature of the effects on welfare. To do this they construct a spatial equilibrium framework where infrastructure facilitates economic activity only through two channels: the reduction of trade costs or the reduction of costs of migration, which translates into a more efficient reallocation of labour, enlarging welfare. This model combined three additional structural parameters which are included as elasticities: one of the good prices to interregional market access, another of migration to differences of real wages, and the third of housing prices to population. The authors find that welfare increased by 2.8%, of which 76% was attributable to reductions in trade costs and 24% was caused by the reallocation of labour. Furthermore, another policy-relevant finding emerges from the framework: trade costs are 20% more prone to changes following are reduction in travel times, hinting that the increase in trade is the driving economic factor behind investments in infrastructure.

4.7 Gibbons & Wu: Airports, Access and Local Economic Performance

Moving away from the traditional methods of transportation of goods, which Morten and Oliveira (2018) have estimated to be the largest factor of economic development in light of transport infrastructure investments, Gibbons and Wu (2020) proceed to analyze the economic effects of improving air travel capacity. The scope of the study is understanding whether

[5] Studied by Atkin (2016), internal immigrants are shown to pay a "caloric tax" to keep eating the same diet that they ate at home while spending more due to the reduced availability of said products.

investments in air travel, which have been commonly prioritized in developing countries over roads or railways due to their smaller scale, are relatively comparable in terms of boost in economic performance to other means of transport. To quantify the causal economic development caused by the opening of new airports, the authors study the opening of 58 new airports in China during the last twenty years.

One key element that the authors had to address is the distinction from the economic development brought by the direct, indirect and induced impact of airport operations to the remaining effects caused by a reduction in transport costs. To resolve this issue and the one relating to non-random allocation patterns the authors take into consideration two groups of counties in new airports that were not built. The differentiating feature of the groups was whether the opening of the new infrastructure did or did not change the travel distance from the closest airport. This was the identifying source of variation of the economic development causally linked to a reduction in transport costs.

This distinction method was made possible by demonstrating that the counties that "incidentally" gained from the investments did not significantly improve their Air Transport Access Index (ATAI), first introduced by Harris (1954), which hints that their gains in transport costs are not to be attributable to policy design.

Their empirical analysis starts by analysing the changes in industrial output, value-added, and GDP, during a medium-run period, from a production function specification which takes the following form:

$$\Delta lny_{it} = \alpha + \beta\Delta \ln(air_{it}) + \gamma x_{it} + \epsilon_{it} \tag{4.7}$$

Where γx_{it} represents the flexible controls that include geographic and socioeconomic characteristics. Meanwhile, the treatment variable is $\Delta \ln(air_{it})$ which encompasses the medium run delta (the data varies from 4 to 9 years) in the expected ATAI. The index, given the pre-existing airport locations, evaluates the travel times between counties and airports, weighing them by their average daily flights. One key element of the index construction is that its findings can be decomposed into two variables which are both of interest: one being the changes in the national availability of airports, and the other being the changes in the availability of airports to a given county.

After identification, the authors proceed to test the validity of their model by exploiting the random variation changes in ATAI in counties that lie in spatial buffer zones. These are areas that after the construction of new airports lie in between the pre-existing and the new airports. To select these counties the authors restrict the study to those counties which lie between 40% and 60% in travel distance between the old and new airports, *de facto* creating a dataset of counties that form rings around new airports.

Gibbons and Wu find that the treatment variable, even accounting for another series of geographic and socioeconomic controls, is stable around an elasticity of 0.20-0.28 for gross output, and 0.25-0.31 for value-added which, given the results permitted the estimation of GDP: 0.19. Interestingly, the GDP estimations show that all gains are attributable to the secondary sector (0.41), while the other sectors do not significantly change. These findings lead to the economic interpretation that every percentage point of higher-than-average ATAI leads to a 4-5 percentage point gain in firm productivity.

4.8 Dappe, Jooste, and Suarez: Port Efficiency, Transport Costs, and Trade

While being responsible for 80% of global trade by volume[6], ports are another piece of infrastructure that is fundamentally understudied in the research field. For this reason, Dappe, Jooste, and Sauron (2017) in their paper study how much port efficiency can increase trade, and consequentially economic performance. This work was also pushed by previous literature findings that linked transport costs and trade flows however, the literature did not manage to identify a constant measure of port efficiency.

To do this the authors used as the target of their study a sample of countries in southern Asia and western Africa, areas which were chosen due to the steep increase in trade in percentage of GDP in ten years leading up to the study. Nonetheless, these countries lagged in trade growth compared to other regions of the world and multiple indicators[7] point to insufficiencies in infrastructure endowment, more specifically in ports, as a primary reason.

To avoid using improper evaluation mechanisms, as was widely done in previous literature on the topic, the authors build a measure of economic efficiency that links port facilities to overall infrastructure output. This measure is constructed using data envelop analysis (DEA), a non-parametric approach that measures the relative efficiency of agents given their input and output data. This is done by constructing an efficiency frontier that is based on the given dataset of

[6] See IMO (2012)
[7] These include the Global Enabling Trade Report (WEF, 2016), World Development Indicators (World Bank, 2023)

firms, without the need to rely on models or functional forms. Given this definition, the subsequent step was choosing the inputs and outputs. To do this, the authors followed a series of indications coming from literature which comprised various measures of port facilities, and labour as inputs, while 20-foot equivalent (TEUs) units were chosen as outputs.

After analysing the country's efficiency rates, a model that linked these and transport costs was used following Fink's (2002) functional form. This is a pricing formula of marginal costs of transporting goods between a pair of ports that are included in its specification: distance, volume, the value of trade, level of containerization, mark-ups, oil prices and, crucially for this study, port efficiency, as well as other controls.

Lastly, the authors proceed to build a gravitational model that connected maritime transport costs with bilateral trade flows that calculate the exports of a country given its dependence factor with the exporter, the elasticity of substitution of goods, and the two countries' GDPs. The subsequent step is to combine Fink's pricing formula into the gravitational model obtaining the desired specification which takes the following form:

$$nx_{jkt} = \alpha_k + \alpha_t + \alpha_j + \beta_1 p_{jkt} + \beta_2 ds_{jt} + \beta_3 gdp_{impt} gdp_{jt} + \varepsilon_t \qquad (4.8)$$

In this equation every parameter is in log form and nx_{jkt} is the exports of country j to country *impt*, ds_{jt} is the vector of minimum maritime distance, p_{jkt} is the unit costs calculated in the pricing formula. Meanwhile, gdp_{impt} and gdp_{jt} represent the two countries GDP's, and $\alpha_k, \alpha_t,$ and α_j represent controls for commodity, time, and country heterogeneity.

Each one of the three study steps reported interesting results: the first phase, estimation shows that, albeit improving over the study period, all the regions find themselves in similar positions and there is great room for improvement. In fact, on a 0-1 scale no group of countries could reach even 0.5 on the efficiency scale. Secondly, improvements in port efficiency of 0.1 on the aforementioned scale lead to a reduction in costs of 2.3%. In the last phase, estimation shows that a 1% reduction in costs increases trade by 0.16%

4.9 Donaldson and Hornbeck: Railroads and American Economic Growth

Before embarking on the study of the Indian railway network, Dave Donaldson (2016), in collaboration with Richard Hornbeck, first studied the causal implications of railway networks on economies in his "*Railroads and American Economic Growth: A Market Access Approach*".

In this paper, the aggregate impact on the agricultural sector following the rapid construction of the US network between 1870 and 1890 is studied.

In contrast to the work before analyzed which used an instrumental variable approach, this paper exploits a reduced form expression of the general equilibrium trade theory which linked the extension of the network with a measure of market access. This measure was built on a database of rail-, road- and waterways during the treatment period and constructed by computing the lowest-cost freight routes, given the extension of the network, in each county, even accounting for intermodal transport.

The main methodological challenge to the correct estimation of the treatment that this article had to tackle was the substantial spillover effects caused by the treatment itself. The solution proposed by the authors was utilizing the changes in market access following the treatment period to calculate direct and indirect impacts on the changes in land values.

This was done thanks to a process of assumption relaxation in comparison to those made in the parent article (Fogel, 1962), obtaining the following specification:

$$ln(V_{ot}) = \beta ln(MA_{ot}) + \delta_o + \delta_{st} + f(x_o, y_0)\delta_t + \epsilon_{ot} \tag{4.9}$$

This regresses the log value of agricultural land in county *o* in year *t* to the log market access measure (MA_{ot})[8], an index of county fixed effects (δ_o), one of the time-variant state-level controls (δ_{st}), and a function of geographic and time effects.

One key feature of the empirical setting is that market access was not determined only by improvements in the railway network. This meant that the gains in the access index could be confronted by the counties that did not receive new railroads during the treatment period, thus resolving the potential issue of endogeneity. They find that market access significantly increased even in countries where no railroads were built within a 40-mile buffer radius, thus demonstrating that rail was commonly used for freight as a medium with other transport means.

After estimating this effect the authors then studied how an increase in market access was capitalized in the agricultural sector by looking at the values of agricultural land. They find that

[8] This was calculated as the summation of all districts by multiplying the destination district population and the relative trade cost (calculated for distance and modes of transport), accounted by the county's bilateral comparative advantages.

removing the improvements brought to the network in the previous twenty years would have reduced agricultural land value by 60.2%, leading to a 3.22% reduction in GNP.

In addition to this exercise, the authors proceeded to calculate the differences in market access in the absence of the railway network at the end of the treatment period, finding a reduction of 80%. Lastly, another counterfactual analysis is brought forward: what if, in the absence of the railway network, waterways or roads could have obtained comparable results in the same timespan? To answer this question existing and proposed routes[9] of these alternative networks were compared. This resulted in a mitigation factor of 13% for the waterways and 21% for the long-distance wagon routes. This demonstrated simultaneously that these networks: did not suffer in their expansion following the railway construction boom, and could not be suited for the functions that the railways had.

4.10 Jedwab, Kerby, and Moradi: History, Path Dependence and Development

One aspect that was touched on by other articles but never directly studied is how transport infrastructure geographically ties economic development to itself, creating a heterogeneous array of consequences to the country's economic geography. One of these consequences is path dependence in spatial distribution and Jedwab, Kerby, and Moradi (2017) study this phenomenon in the setting of colonial Kenya. They use the construction of the railway lines, which occurred for military and strategic interests of the British Empire at the end of the 19th century, as a source of an exogenous shock to the spatial equilibrium, which permanently disrupted it.

In fact, following the construction, a series of settlements sprouted along these railway lines, first as operational facilities for European farmers who started utilizing the surrounding land, and secondly as true urban centres thanks to the influx of Asian settlers who were mainly traders. However, this newfound physical and human equilibrium was then remodified in the post-independence period when both of the aforementioned groups mainly left the country and the railway network fell into disuse due to mismanagement, thus giving the following treatment period: 1901-1962.

The first step in the analysis was quantifying the displacement effect on the different communities caused by the introduction of the railway. This was done by comparing the

---

[9] Obtained from the database of the parent paper *Railroads and Economic Growth* from Fogel (1962)

populations of connected and non-connected areas at the height of the railway use, so just before independence in 1962, to just before the railway construction in 1901, knowing that the non-African population were close to zero at the start of the treatment and native population patterns did not coincide with the railway routes. The specification uses, for intercultural and intertemporal comparisons' sake, standardized *z* values and has the following form:

$$zPop_{l,62} = \alpha + \beta Rail_{l,62} + \omega_p + \xi X_l + \upsilon_{l,62} \quad (4.10)$$

In this equation *Pop* is the standard score of urban, native, European, and Asian populations in location *l* in 1962, *Rail* is a dummy which encompasses the bandwidth distance from *l* to the railway, ω_p and X_l are a series of province and location fixed-effects.

The authors found that, in accordance with previous literature, the investments in railroads had a long-term positive impact on overall economic development, especially in promoting: urban growth for skilled individuals, which majorly affected non-native populations, and agricultural production for European farmers.

Once having assessed this, the setting provided the possibility of understanding the channels which brought path dependence by confronting the economic performance of the before-mentioned urban locations *l* in the period after Kenya's independence (1962-2009). This exercise used a similar specification[10] to (4.10), with the dependent variable being only the urban population and including all ethnic groups as independent variables and the variable of interest being still the rail connectivity dummy. To understand the channels at work a series of additional controls proxied the proposed channels which were: institutional persistence, changes in transport technology, sunk investments, and spatial coordination failures.

Interestingly for my work, among the channels at work, the findings show that transport infrastructure, both rail- and roadways, were not a cause of path dependence in the spatial equilibrium *per se* since spatial equilibrium is not directly affected by the demise of the railway network. However, it had still significant effects on human capital's spatial coordination failures, meaning that the higher population densities of treated cities naturally attracted proportionally more population than economic factors might indicate. Another channel that is

[10] In fact, all the province- and location-level fixed effects remained the same

equally at work is the sunk costs in other types of public and private infrastructure that were present at the moment of independence.

4.11 Heblich, Redding & Sturm: The Making of Modern Metropolis

The last paper I will analyse exploits the technological innovation of transport infrastructure as the main vector of urban reallocation and repurposing. In their paper "*The Making of the Modern Metropolis*", Heblich, Redding, and Sturm (2020) investigate how the introduction of the steam locomotive in the urban setting of 19th-century London drastically changed the spatial distribution of the city, allowing for more distant commuters, and a specialization of the centre toward a solely commercial area, thus increasing the possibilities for population growth.

To do this, the authors introduce a structural equation model characterized by commuting flows through a gravitational perspective to determine commuting costs. With these, they could calculate the historical unobserved values of employment location and commuting patterns before the treatment took place. And, to do this, they first had to resolve the endogeneity issue by including parish-level time-invariant controls to account for the railway companies' intent of targeting more prosperous and populous areas. However, it is important to remark that, to be applicable, the model requires the assumptions of commuter gravity, market clearing, and proportional rents to income[11] to be made.

The findings show that there is a sharp decline in the City population with an increase in non-central parishes that received a railway station in a range between 11% and 133%, with the biggest estimates coming from those lying in a 10 to 20-kilometre radius. Interestingly, the authors also find discontinuities in parish population trends in the years right after the opening of the station, indicating that the increase in population is directly correlated with it.

After confirming that the railroads caused a distinct shift in the urban's spatial equilibrium, the authors proceed with a counterfactual exercise which aims at understanding what would have happened if this commuting technology did not exist. To do this they exploit the recursive nature of the model to estimate the unobserved changes in residence, employment, and property values/rents without the need to make ulterior assumptions regarding the other variables in the

---

[11] Gravity in commuting relates to the willingness to face the (physical and monetary) costs of commuting to enter a bigger market (in this case the commercial City of London), market clearing relates to a conjoined set of assumptions that dictate the rules of the market such as rationality of agents, perfect competition, and price flexibility. Meanwhile, that rents are constantly proportional to income means that residential rents are constant to residential income and commercial floor spaces are constant to workplace income.

model. This exercise resulted in a reduction in the Greater London population of 33% and property values of 22%. Furthermore, they find that there would have been a reduction in inter-borough commuters from 370'000 to less than 60'000.

To conclude, the authors carry on a reduced[12] cost-benefit analysis of the construction of the railway network using the historical estimates of the building costs of said network and comparing those to the changes in the net present value of floor space and city-wide GDP. Furthermore, two estimates coming from the previous model are brought forward regarding the elasticities of the supply of floor space and agglomeration forces. Obtaining that, even with the upper-bound discount value of 5%, the railway network provided benefits for 14.10 and 24.31 times the construction cost for the overall value of floor space and revenue respectively.

5. Publication Bias

As is the case with different fields of economic literature (Christensen and Miguel, 2018), it is critical to understand whether transport infrastructure research is afflicted by publication bias. With publication bias, I refer to the preference of researchers, editors, and reviewers to publish statistically significant results, limiting the publishing of studies that obtain insignificant or even opposite results to what the rest of the literature has discovered (Andrews and Kasy, 2019). This problem causes the literary corpus to be skewed in its findings, losing some credibility in the eyes of policy-makers and practitioners. For this reason, it is crucial to understand whether this field of research suffers from this problem, and if so, finding some possible solutions is the subsequent step.

One method to recognize the presence of said bias was first introduced by Stanley and Doucouliagos (2012), called the FAT-PET-PEESE approach: this stands for the three steps necessary to conduct the assessment. First comes the *Funnel Asymmetry Test*, which relates the magnitude of the findings in the literature with their standard errors. The two are put together in the OLS regression equation 5.1 where $\hat{Q}_{es}$ is the effect *e* of a study *s* and *SE* are the standard errors of the studies. An unbiased group of publications would accept the null hypothesis that any kind of correlation between the two.

[12] Reduced since the costs regarding operations, market structure, and externalities are not accounted for.

$$\hat{Q}_{es} = \beta_0 + \beta_1 SE_{es} + \varepsilon_{es} \tag{5.1}$$

The second step is called the *Precision-Effect Test*, which is focused on finding whether there is a non-zero effect of publication bias on the analyzed set of findings. This is done by dividing the above equation by its standard error:

$$t_{es} = \beta_0 \frac{1}{SE_{es}} + \beta_1 + \epsilon_{es} \tag{5.2}$$

where ϵ_{es} is equal to $\frac{\varepsilon_{es}}{SE_{es}}$ and t_{es} is coincidentally $\frac{\hat{Q}_{es}}{SE_{es}}$. Finding a zero effect of publication bias means being able to accept H_0. Lastly, the final stage of the approach is defined as *Precision-Effect Estimate with Standard Errors* and works to define what is the significance threshold of publication bias. This uses the same equation (5.1), excluding the use of variance SE_{es}^2 instead of the standard errors, and uses the same transformation in (5.2) to obtain the following:

$$t_{es} = \beta_0 \frac{1}{SE_{es}} + \beta_1 SE_{es} + \epsilon_{es} \tag{5.3}$$

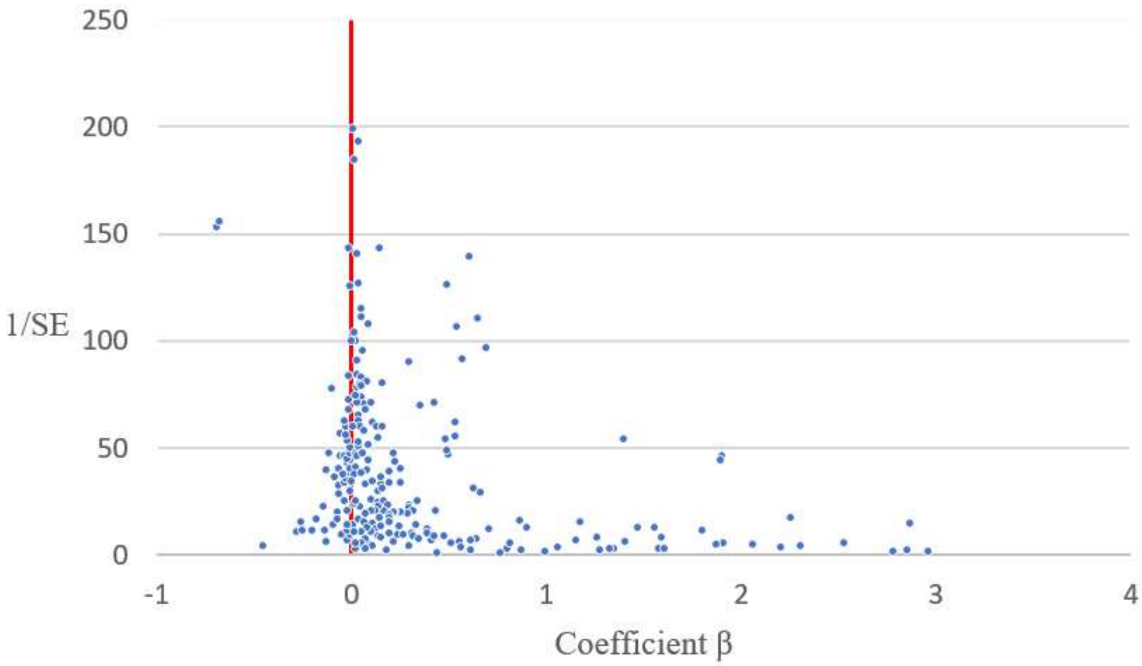
Given this equation a set of additional steps are required to remove the issue of heteroskedasticity of the sample to obtain the following final regression that includes a set of moderators indicated with M_{ke}:

$$t_e = \beta_0 \frac{1}{SE_e} + \beta_1 + \sum \beta_k \frac{M_{ke}}{SE_e} + u_e \tag{5.4}$$

In this regression failing to accept $H_0: \beta_1 = 0$ would mean acknowledging the presence of publication bias.

Given this last regression, I borrow Foster et al. (2023b) set of ninety transport-related studies published in the thirty years which encompass 223 regressions to conduct a comprehensive review of the historic state of publication bias in the field. In their work, they find that the presence of publication bias is estimated to be statistically significant at a 1% confidence level. This is further confirmed by the visual representation of the estimates provided in Figure X. It is important to note that one of the moderator variables included in their work in equation (5.4) was a factor of time and the regression has shown that research in the transport sector has produced larger estimates over time.

Figure 1: Funnel Asymmetry Plot



Source: Foster et al. (2023b)

Having assessed the presence of bias, it is in the academic world's interest as a whole to design policies and methods to tackle this. Some possible methods have been put in place by other fields of research to reduce this problem, but focusing on the economic research environment, in the last decade there has been a notable increase in awareness of transparency and reproducibility efforts (Christensen and Miguel, 2018). This is being achieved with a progressive popularization of: result-blind reviews for reviewers[13], pre-analysis plans for researchers (Olken, 2015), and open-data settings for replication papers. Furthermore, during the conduction of literature reviews and meta-analysis, another effort consists of the inclusion of a combination of non-published articles, working papers, pre-prints and reports.

[13] see the *American Journal of Health Economics*

6. Conclusion

This work has aimed to conduct a synthesis of the main studies in the transport economics literature in the last 15 years, providing insights into the theoretical and econometric innovations, as well as the findings that disrupted the field of research.

The first finding of this review, in terms of importance, is the assurance that infrastructure, albeit often through complex application mechanisms and unforeseeable localized negative consequences, is a vital element in the development process of any country. In particular, developing countries are found to better gain from these investments which, on average, show 44.7% larger elasticities (Foster et al., 2023b). Nonetheless, the before-analyzed findings should take into account that the whole literary corpus suffers from publication bias.

Closely related to this, the work shows that, even if a myriad of channels is simultaneously at work, the main economic force that boosts development following improvements in transport infrastructure endowment is the reduction in trade costs, which allows for an increase in trade. Meanwhile, other channels such as the reduction in commuting and migration costs tend to have a smaller magnitude of impact.

Another result of the work is trying to identify how elasticities could be translated into policy recommendations. As previously seen, there are multiple instances where the policymaker influences the future investment's rate of return. However, this translation is theoretically unfeasible due to the unavailability of internationally coherent data on the stocks of infrastructure as discussed by Fay et al. (2019). This shortcoming has proven to be one of the biggest obstacles that the field has to tackle to obtain more relevant policy suggestions.

Furthermore, the work has shown that researchers tend to favour unique settings that permit an easement in data and/or methodological tractability. This brings however two consequences: first it hinders the possibilities for long-term reproducibility efforts of the studies, in turn, facilitating the perseverance of publication bias. Secondly, it confirms the thesis of Redding and Rossi-Hansberg (2017), which states that in spatial equilibrium's theoretical models, favouring analytical tractability means impoverishing the model itself, leading to unique models that are hardly comparable.

One other finding of the conducted survey is that there is ever-increasing attention given to the externalities that transport infrastructure have outside of household- and firm-level transport costs, especially concerning carbon emission (Couture, Duranton & Turner, 2018) and congestion (Parry, Walls & Harrington, 2007).

Lastly, the literature survey demonstrated that researchers tended to favour the study of roads and railways, however, thanks to this new wave of studies there has been a newfound interest toward airports, water ports, and inland ports (Foster et al., 2023a).

References

ANAS, A., and LIU, Y., 2007. The geography of air passenger traffic and city size: Evidence from U.S. metropolitan areas. *Journal of Regional Science*, 47(3), 457-475.

ANDREWS, I., and KASY, M., 2019. Identification of and Correction for Publication Bias. *The American Economic Review*, *109*(8), 2766–2794.

ASCHAUER, D. A., 1989. Is public expenditure productive? *Journal of Monetary Economics*, *23*(2), 177–200.

ASHER, S., and NOVOSAD, P., 2020. Rural Roads and Local Economic Development. *American Economic Review*, *110*(3), 797–823.

ATKIN, D., 2016. The Caloric Costs of Culture: Evidence from Indian Migrants. *The American Economic Review*, *106*(4), 1144–1181.

BAUM-SNOW, N., 2007. Did Highways Cause Suburbanization? *Quarterly Journal of Economics*, *122*(2), 775–805.

BAUM-SNOW, N., et al., 2017. Roads, Railroads, and Decentralization of Chinese Cities. *The Review of Economics and Statistics*, *99*(3), 435–448.

BERG, C. N., et al., 2017. Transport Policies and Development. *The Journal of Development Studies*, *53*(4), 465–480.

COUTURE, V., DURANTON, G., and TURNER, M. C., 2018. Speed. *The Review of Economics and Statistics*, *100*(4), 725–739.

CHANDRA, A., and THOMPSON, E., 2000. Does public infrastructure affect economic activity? *Regional Science and Urban Economics*, *30*(4), 457–490.

CHRISTENSEN, G., and MIGUEL, E. C., 2018. Transparency, Reproducibility, and the Credibility of Economics Research. *Journal of Economic Literature*, *56*(3), 920–980.

DAPPE, M.H., JOOSTE C., and SUAREZ-ALEMAN A, 2017. How Does Port Efficiency Affect Maritime Transport Costs and Trade? Evidence from Indian and Western Pacific Ocean Countries. *Policy Research Working Papers,* 8204. © World Bank, Washington, DC.

DENG, T., 2013. Impacts of Transport Infrastructure on Productivity and Economic Growth: Recent Advances and Research Challenges. *Transport Reviews*, *33*(6), 686–699.

DONALDSON, D., and HORNBECK, R., 2016. Railroads and American Economic Growth: A "Market Access" Approach *. *Quarterly Journal of Economics*, *131*(2), 799–858.

DONALDSON, D., 2018. Railroads of the Raj: Estimating the Impact of Transportation Infrastructure. *The American Economic Review*, *108*(4–5), 899–934.

DURANTON, G., and TURNER, M. C., 2012. Urban Growth and Transportation. *The Review of Economic Studies*, *79*(4), 1407–1440.

EATON, J., and KORTUM, S., 2002. Technology, Geography, and Trade. *Econometrica*, *70*(5), 1741–1779.

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

FAY, M., et al., 2019. Hitting the trillion mark: a look at how much countries are spending on infrastructure. *Policy Research Working Paper*, 8730. © World Bank, Washington, D.C.

FINK, C., 2002. Trade in International Maritime Services: How Much Does Policy Matter? *The World Bank Economic Review*, *16*(1), 81–108.

FOGEL, R. W., 1962. A Quantitative Approach to the Study of Railroads in American Economic Growth: A Report of Some Preliminary Findings. *The Journal of Economic History*, *22*(2), 163–197.

FOSTER, V., et al., 2023. The Impact of Infrastructure on Development Outcomes: A Qualitative Review of Four Decades of Literature. *Policy Research Working Papers,* 10343. © World Bank, Washington, DC.

FOSTER, V., et al., 2023. The Impact of Infrastructure on Development Outcomes: A Meta-Analysis. *Policy Research Working Papers*, 10350. © World Bank, Washington, DC.

GARCIA-LÓPEZ, M., HOLL, A., and VILADECANS-MARSAL, E., 2015. Suburbanization and highways in Spain when the Romans and the Bourbons still shape its cities. *Journal of Urban Economics*, *85*, 52–67.

GHANI, E., GOSWAMI, A. G., and KERR, W. R., 2016. 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., and WU, W., 2020. Airports, access and local economic performance: evidence from China. *Journal of Economic Geography*, *20*(4), 903–937.

HARRIS, C. D., 1954. The market as a factor in the localization of industry in the United States. *Annals of the Association of American Geographers*, 44, 315–348.

HEBLICH, S., REDDING, S. J., and STURM, D. M., 2020. The Making of the Modern Metropolis: Evidence from London. *Quarterly Journal of Economics*, *135*(4), 2059–2133.

IMBENS, G. W., and LEMIEUX, T., 2008. Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, *142*(2), 615–635.

IMO (International Maritime Organization), 2012. International Shipping Facts and Figures: Information Resources on Trade, Safety, Security and Environment. London: Maritime Knowledge Center.

ITF (International Transport Forum), 2021. *ITF Transport Outlook 2021*, Paris: OECD Publishing.

JEDWAB, R., KERBY, E., and MORADI, A., 2017. History, Path Dependence and Development: Evidence from Colonial Railways, Settlers and Cities In Kenya. *The Economic Journal*, *127*(603), 1467–1494.

MORTEN, M., and OLIVEIRA, J., 2018. The effects of Roads on Trade and Migration: Evidence from a Planned Capital City. *NBER Working Paper, 22158*, 1-64.

OLKEN, B. A., 2015. Promises and Perils of Pre-analysis Plans. *Journal of Economic Perspectives,* 29(3), 61-80.

PARRY, I. W., WALLS, M., and HARRINGTON, W., 2007. Automobile Externalities and Policies. *Journal of Economic Literature*, *45*(2), 373–399.

REDDING, S. J., and TURNER, M. C., 2015. Transportation Costs and the Spatial Organization of Economic Activity. *Handbook of Regional and Urban Economics, 5,* (1339-1398).

REDDING, S. J., and ROSSI-HANSBERG, E., 2017. Quantitative Spatial Economics. *Annual Review of Economics*, *9*(1), 21–58.

STANLEY, T. D., and DOUCOULIAGOS, H., 2012. *Meta-regression analysis in economics and business*. Routledge.

WEF (World Economic Forum) and Global Alliance for Trade Facilitation, 2016. The Global Enabling Trade Report. Cologny: World Economic Forum

WENZ, L., et al., 2020. Road to glory or highway to hell? Global road access and climate change mitigation. *Environmental Research Letters*, 15(7), 075010.

WORLD BANK, 2023. World Development Indicators. Washington D.C.: World Bank Publishing.