A Pseudo - Panel Approach to Estimating Dynamic Effects of Road Infrastructure on Firm Performance in a Developing Country Context

Samira Barzin[*][a]

Imperial College London

Sabine D'Costa[b]

The University of Westminster

Daniel J. Graham[a]

Imperial College London

January 2018

*Corresponding Author: s.barzin12@imperial.ac.uk

[a]Department of Civil and Environmental Engineering, Centre for Transport Studies, Imperial College London, London SW7 2AZ, UK

[b]Department of Economics and Quantitative Methods, The University of Westminster, London NW1 5LS, UK

Abstract

To overcome the absence of true firm-level data, we provide evidence that the use of pseudo-panels based on aggregated data can correctly identify production function parameters. We construct a pseudo-panel of Colombian manufacturing firms for the years of 2000 to 2009 to study the effects of transportation infrastructure on firm performance in a developing country and find elasticities of output with respect to road infrastructure ranging from 0.13 to 0.15 per cent. This confirms that roads are important for private output growth and, as our results are larger than those reported in the literature for developed countries, that transportation infrastructure is relatively more important for the economy of developing countries. We also identify a one-year time lag with which firms' outputs react to road stock changes. This could be indicative of firms requiring time to adjust their production to road changes. We furthermore identify that the effect of road infrastructure is particularly large for heavy manufacturing industries. Moreover, we investigate the regional heterogeneity of the role of transportation infrastructure for firms' output growth. Our results are robust to different econometric concerns. We additionally provide Monte Carlo simulations to support the validity of pseudo-panels in the context of firm-level data.

I   Introduction

Transportation infrastructure is a crucial component to economic growth (Crafts, 2009; Tripathi and Gautam, 2010). However, while the majority of developed countries possess relatively dense transport networks, developing countries often suffer from a low road stock and underinvestment in infrastructure. This is a particular problem for Latin American countries for which infrastructure stocks have noticeably fallen behind the industrialised Western and East Asian countries since the 1970s. Additionally, with an average infrastructure spending of around 1 per cent of GDP across Latin America, infrastructure investments have barely grown in the 2000s (Calderón and Servén, 2010).

Colombia has recently launched an immense road transportation programme consisting of 40 public-private partnerships to build 8,000 kilometres of highway road infrastructure by the year 2020. The main goal of this project, which is estimated to cost around 25bn US Dollars, is to connect the main economic centres of the country with each other and to the ports of the Atlantic and Pacific oceans through interconnected four lane highways. This vast programme also includes the "Highway to Prosperity" project situated in the Northwest of the country. Deemed currently as the most extensive transportation project globally with an expected cost of 7.2bn US Dollars, it aims at establishing north to south and east to west transportation links. Additionally, with a recently signed free trade agreement between Colombia and the US, the road investment project is expected to increase trade volumes and furthermore aid economic development[1].

While one can only forecast the economic benefits accruing to the Colombian economy from these extensive projects, we provide an insight into the relationship between road infrastructure and the Colombian economy by conducting an ex-post evaluation for the years 2000 to 2009. We use a pseudo-panel of Colombian manufacturing firms for the analysis which relies on data from the Annual Manufacturing Survey conducted by the Colombian statistical authority DANE. Estimating the effects of road infrastructure on the production of manufacturing firms, we find that while current highway infrastructure appears insignificant across all specifications, lagged highway stock affects output growth positively and significantly. Our results indicate that a growth in transportation infrastructure of 10 per cent, results in manufacturing output growth of 1.31 to 1.53 per cent in the subsequent year for the whole sample. Results are significantly larger for heavy industries and are heterogeneous across the regions. These results suggest that firms' production processes require time

---

[1]See reports by Agencia Nacional de Infraestructura de Colombia/ANI, 2013; Buendia and Gargan, 2012; Departamento Nacional de Planeacion de Colombia/DNP, 2010.

to adjust to highway expansions. The identified elasticities furthermore indicate that the returns from highway expansions on firms are notably larger for developing countries relatively to firms in developed countries with extensive transportation networks.

This paper is structured as follows: We provide an overview of the related literature in Section 2 and discuss the data sources in Section 3. We outline the econometric model for the analysis and provide a discussion of the pseudo-panel methodology and the associated Monte Carlo simulations in Section 4. Section 5 analyses the results and Section 6 concludes.

II   An Overview of the Context and the Related Literature

Transportation infrastructure affects economic variables through various channels[2]. Direct benefits from improvements of transportation infrastructure arise from increases in connectivity due to reductions in travel times and travel costs for both goods and people. This results in logistic benefits for the transportation of intermediate input and final output goods, and also allows for faster and less costly commuting of employees (Gimenez-Nadal and Molina, 2014). These benefits will have direct positive effects on the level of productivity of firms.

Additional benefits arise from changes in agglomeration economies and effective density stemming from transportation cost reductions (Graham, 2007). These encompass the sharing of resources across larger geographical space, more efficient matching between employers and employees, and across business partners, and increased information exchange through knowledge sharing and faster learning (Duranton and Puga, 2004). Further, lower transportation costs will enable firms to reach distant markets faster and at lower costs. Hence, transportation improvements may also result in market expansions and increased levels of competition. A heightened level of competition will consequently force less productive firms out of the market and will simultaneously further raise the pressure on surviving firms to increase their productivity level so that the overall degree of average productivity will increase (Baldwin and Okubo, 2006; Melitz, 2003). This may consequently encourage specialization across firms as comparative advantages become relatively more important to remain in a more competitive market (Bougheas et al., 2000). Furthermore, changes in transport costs can also affect a firm's input choices. If transport cost reductions result in changes in the relative prices of the intermediate inputs, it may be optimal for the firm to change its input factor mix (Holl, 2006). Additionally, if transport cost reductions yield price reductions of the final good,

---

[2]See Venables *et al.* (2014) for an extensive review of the productivity effects of transportation infrastructure

increased demand for the firm's output could be generated (Lahr et al., 2005). Moreover, as trucks often depreciate at a lower rate on better quality roads, better transportation infrastructure also allows for a longer lifespan of the existing capital stock (Barnes and Langworthy, 2003). Further, private capital and transportation infrastructure are often considered to be complements. Transportation infrastructure can make a region a more attractive location and hence encourage private investments (Crafts, 2009). This will in turn create new demand for labour and generally foster economic growth.

The groundwork for the empirical research on the economic effects of road infrastructure was provided by Aschauer's (1989) work on the economic effects of public infrastructure in the US. He includes local, state and federal capital stock consisting of structures and equipment into his infrastructure measure to estimate the economic effects of public infrastructure. Aschauer estimates an aggregate production function that includes public infrastructure and identifies an elasticity of 0.35 of aggregate production with respect to public infrastructure spending. Furthermore, he finds that 55 per cent of this public infrastructure effect arise from energy and transportation infrastructure. Subsequent work by Fernald (1999) singles out the role of transportation infrastructure and its effects on productivity in the US. He includes transportation infrastructure as an additional input factor in a production function and finds effects of very similar magnitude to those identified by Aschauer (1989).

While this early literature focusing on the role of transportation infrastructure on the economy laid the empirical foundation for future work in this field, it has since been heavily criticized on the basis of various econometric issues. As a response, more recent papers rely on more complex estimation strategies to prevent biases from endogeneity issues, e.g., due to endogenous road placement in areas where output growth is expected. While the literature has moved away from solely focusing on the effects of transportation on productivity or output growth, it provides a more thorough insight into the role of transportation on the economy as a system. Duranton and Turner (2012) estimate the effect of US interstate highways on city employment. Their identification strategy relies on an instrumental variable approach that employs a historical highway plan and railroad map as exogenous factors as determinants of current highways but not of current employment growth. Their results show that a 10 percentage point increase in a city's interstate highway stock yields a 1.5 percentage point increase in its employment over the subsequent 20 years. Holl (2012) investigates the influence of transportation infrastructure on firm-level productivity through its effects on market potential in Spain. She constructs a firm's market potential based on travel times, which in turn depend

strongly on the existing road network. To rule out any endogeneity biases, she relies on historical data to construct instrumental variables for current market access. Her reported estimates of the growth of market access on output growth range from 0.041 to 0.074. While instruments based on historical data have become a commonly employed method of preventing endogeneity biases, Faber (2014) uses an alternative identification method. He estimates the effect of the Chinese highway network on the spatial distribution of economic activity by researching peripheral towns that were solely connected to the network as they were geographically located between targeted cities. In order to further prevent any endogeneity biases, he constructs two instrumental variables that are based on hypothetical least costly road paths for the highway's construction. His results suggest that the construction of the highway system resulted in significant reductions of GDP growth and industrial output in the peripheral regions. It further allowed for trade cost reductions that consequently shifted economic activity from peripheral regions towards cities. While the majority of the research to date has focused on developed countries, infrastructure has been identified as an important driver for economic development in developing countries and insufficient infrastructure as a crucial impediment for development. This notion is supported by findings of The World Bank which regularly surveys firms and entrepreneurs doing business in developing countries for their World Bank Investment Climate report. The report has identified that 20 per cent of the surveyed sample in East Asia and Pacific and 55 per cent in the Middle East, North Africa and Latin America find insufficient electricity, telecommunications and transport infrastructure as a severe obstacle to doing business (The World Bank, 2004).

Calderón and Servén (2004a) estimate the effects of infrastructure on GDP using a large panel of 120 developed and developing countries from 1960 to 2000. They use an infrastructure index, consisting of both infrastructure quantity and quality measures, and find results indicating that GDP growth is positively influenced by all included infrastructure factors. Focusing the analysis on Latin America, Calderón and Servén (2004b) identify positive and significant contributions of telecommunications, electricity and transportation infrastructure to per worker GDP growth. Additionally, they show that the marginal products of all three infrastructure measures included significantly exceed those of non-infrastructure capital. They also find that the output gap between Latin American and East Asian countries throughout the 1980s and 1990s can largely be attributed to the different stocks of infrastructure.

Focusing on the regional level in India, Lall (2007) uses a pooled data set of Indian states and finds that transportation and communication infrastructure significantly and positively affect state-level

output growth. Additionally, he identifies that the influence of transportation and communication on economic growth is larger in lagging states.

Highlighting the role of different micro- and macroeconomic factors for the export propensities of Indonesian manufacturing firms, Rodríguez-Pose et al. (2013) identify the effect of transportation infrastructure as a particularly crucial factor determining firms' export patterns. The authors further show that both the road infrastructure of the region where the firm is located and that of neighbouring regions influence firms' export patterns.

Recently, a burgeoning body of research has started to investigate the relationship between firm performance and infrastructure with notable work provided by Ghani et al. (2016) for India and Banerjee et al. (2012) for China. While this work opens the discussion on effects of transportation on firms in developing countries, it predominantly relies on data from China and India which both have established manufacturing sector censuses.

While, as outlined above, there exists some research on the effects of transportation infrastructure on economic growth for developing countries, to date this literature remains limited. Additionally, given the absence of reliable firm level (panel) data for most developing countries, this is particularly the case for research using firm-level data. As economies at different stages of development differ largely in their economic structure, it cannot be assumed that conclusions drawn from research on developed countries also hold for developing countries[3]. Furthermore, the road stock and density are also notably different in developed and developing countries: while the former often have well-developed and dense road networks, the latter often exhibit limited transportation infrastructure and low road densities.

This paper contributes to the literature by taking a microeconomic approach in a developing country context. We use aggregated Colombian firm data and combine it with transportation data to estimate the effects of the highway network on firm-level output. This paper particularly relates to recent work by Duranton (2015) and Blyde (2013) who focus on the effects of roads on trade patterns in Colombia. While Duranton focuses on the effects of within and intercity highway stock on exports, Blyde focuses on the effects of road quality improvements on export patterns. Both of the above papers investigate the relationship between the Colombian economy and transportation infrastructure, however as they exclusively focus on trade, we extend this research by using aggregated firm data to focus on the role of roads on output growth. Additionally, our paper contributes to the literature focusing on investigating the impacts of transportation infrastructure in

---

[3]See for example Hansen (1965)

the absence of reliable panel data. While Storeygard's (2016) work on road infrastructure and city growth for 15 African countries for example proposes the use of satellite lights data as a proxy for income when reported economic data is not available, we highlight the construction of pseudo-panels based on aggregated firm data to research firm patterns as an alternative approach.

III   Data

III.I   A Brief Overview of the Colombian Transportation Sector

Following the 1991 constitutional reform in Colombia and its subsequent changes to the political system throughout the 1990s, the years of 2000 to 2009 represent a time of relative political stability for Colombia. During the majority of this decade, Colombia was governed by the same president and Minister of Transport. This allowed for a consistent policy design during this decade. At the beginning of the 2000s, globally and regionally, Colombia exhibited one of the largest transportation infrastructure gaps relative to its per capita GDP income levels (Calderón and Servén, 2004b). However, the increased focus on Free Trade Agreements during the decade shifted the policy attention on improving the Colombian transportation network. Overall, the policy focus of the transport policy makers during the first decade of the 2000s were travel cost reductions, sustainable regional economic growth, improvements in regional integration, and increases in regional competitiveness (Departamento Nacional de Planeación/DNP, 2002, 2006). Further, a focus was set on improving urban transportation projects with the aim of reducing poverty and reaching employment and equality goals[4]. During the decade investigated, investments in roads represented 32.6 per cent of total public investment, equating to 0.5 per cent of GDP. While 26 per cent of the overall road infrastructure budget were spent on road construction, the majority of 68 per cent were made available for road maintenance projects. Further, within the road infrastructure budget, primary roads received the largest budget of 75 per cent, while projects on secondary and tertiary roads were allocated 7 and 18 per cent of the total budget respectively (Nieto-Parra et al., 2013).

The overall budget on roads remained relatively constant throughout the years of this analysis, the relative distribution of the funds across the regions however exhibited disparity. During the first part of the decade, the focus was set particularly on regions located in the centre and in the East which were allocated the largest absolute amounts of the budget. Further, regions on

---

[4]The analysis of urban transportation projects is beyond the scope of this paper
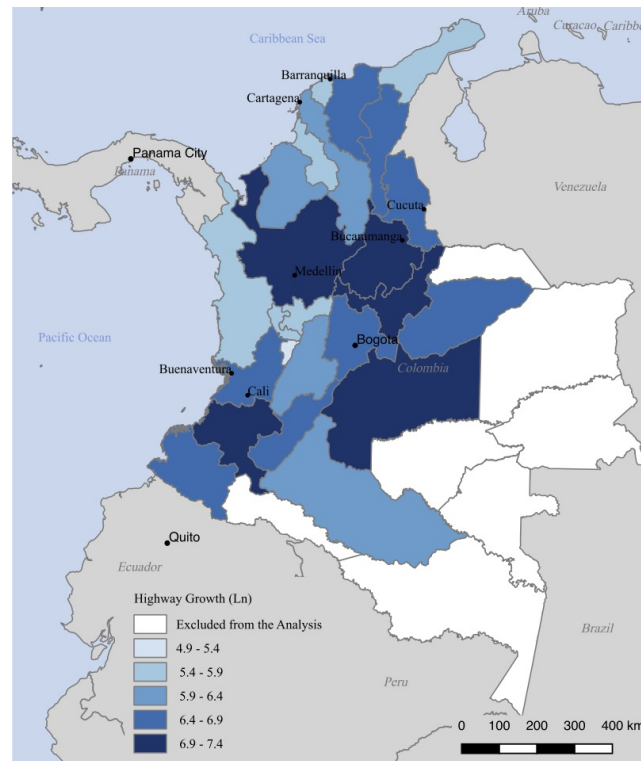
Figure 1: Average Highway Growth (2000 - 2009). Source: Own elaboration on the basis of the road infrastructure data provided to the authors by INVÌAS; No consistent road data was available for the islands of San Andrés and Providencia, so that these could not be included in this study.

the Atlantic coast experienced the largest relative increase in road investments during this time. An additional budget for road infrastructure projects was allocated to the regions according to their relative national economic importance where the regions of Antioquia (located on the Atlantic coast), Valle del Cauca (located on the Pacific coast) and the capital region of Bogotá, D.C. received relatively higher shares of the overall budget. In the second part of the decade, the focus shifted predominantly to regions located on the Atlantic coast and across the Central and Eastern part of the country. A relative increase in allocated funds could also be observed for the Western regions (Nieto-Parra et al., 2013).

The analysis of this paper focuses on the regional changes in the primary highway network during the years of 2000 to 2009. The main source of data used to measure roads is annual data on the Colombian highway stock (km) per Colombian region. This data was provided by the National Roads Institute of Colombia INVÍAS. The advantage of this data set lays in the fact that highways are measured in physical units rather than monetary units which allows for a reduction in measurement errors, and also decreases the effect of possible inefficiencies in the policy implementation process on

the data. Figure 1 provides a graphical representation of regional highway growth across the years included in this analysis.

The majority of the regions of Arauca, Amazonas, Guainía, Guaviare, Putumayo are not covered by the national highway network and are hence excluded. No consistent highway data could be obtained for the region of Vichada; thus, this region was additionally dropped from the sample. Annual data covering the years 2000 to 2009 were used for the analysis. Given that the highway network of the capital region of Bogotá, D.C. is consolidated with the highway network of the surrounding region of Cundinamarca, Bogotá, D.C. is treated as part of Cundinamarca in the context of its highway network.

III.II   Manufacturing Data

The information on aggregated firm data was taken from the annually conducted Colombian Manufacturing census (Encuesta Anual Manufacturera) and is used to obtain information on the output and input factors of the manufacturing sector. This data set covers all manufacturing firms with a minimum of 10 employees and provides information on, inter alia, output, capital stock, employment, inventories, raw materials usage, electricity usage, and investments. The data were provided to us aggregated by sector. The sectors were categorised into three-digit industrial identifiers for each Colombian region (departamento). according to the International Standard Industrial Classification system (ISIC), Rev. 3. We employ the information on the number of firms included in each three-digit industrial sector-departamento pair to generate a pseudo - panel encompassing 4,023 observations[5]from the underlying 74,657 firms. In the constructed pseudo-panel, each observation represents an average firm for a given industry in each region in a given year[6] where the average firm in this context is defined as containing the mean variable values across each industrial sector-departamento pairing. The pseudo-panel then relies on the identified mean variables of each individual cohort, defined here as a region-industry pairing, for the subsequent analysis. The constructed pseudo-panel contains 480 unique cohorts of which 344 are observed for all ten years of this study, the remainder of cohorts are observed for subperiods resulting in an unbalanced pseudo-panel of 4,023 observations across the study period from 2000 to 2009. The compiled data set includes data from the capital district of Bogotá, D.C. and 24 of the 32 Colombian regions. The regions of Casanare, Vaupes and the island state of San Andrés y Providencia were excluded due to insufficient economic data. In order to exclude that the effect of highway expansions

---

[5]Further details on the pseudo-panel methodology can be found in Section 4
[6]Additional descriptive statistics on the regional level can be found in Appendix A

Table 1: Descriptive Statistics

Variable	Mean	Standard Deviation
Output	14,466,326	51,730,405
Capital	10,682,254	38,679,786
Employment	49.44	60.34
Energy	2,922,896	11,451,002
Materials	8,593,099	43,082,174
Highways	793.52	401.39

on output growth is driven by selection and competition mechanisms across firms, we conduct preliminary tests on the data. The results do not reveal any evidence for strong effects of highway growth on the average firm size and the number of firms within regions, hence allowing us to conclude that the effect of highways on output growth is not determined through firm selection[7].

Information on the labour market was taken from the Encuesta Nacional de Hogares and the Gran Encuesta de Hogares for the years 1996 to 2000 and 2001 to 2010 respectively. Both of these labour market surveys provide, among others, information on the working age population, unemployment rates, and the amount of employed for each region. Data on urban and municipal population was taken from the 2005 General Census. The Encuesta Anual Manufacturera, the Gran Encuesta de Hogares, the Encuesta Nacional de Hogares and information from the 2005 General Census have been obtained through the Departamento Administrativo Nacional de Estadística DANE.

Output, capital and raw materials were provided as measured in thousands of Colombian pesos. In order to compute approximate physical quantities of these variables, output is deflated using the producer price indices at the two-digit ISIC level, capital is deflated with the producer price index for the manufacturing of machinery and equipment, and raw materials are deflated by the annual average manufacturing producer price index. Adjusting these variables to both inflation and price differences across industries allows for the approximate identification of deflated physical units from the variables that were measured in monetary units. These are listed in Table 1. Energy, labour and the highway infrastructure stock are measured in physical units. Energy is measured in KWH, labour measures total permanent employment and highway infrastructure is measured by kilometres of highway per Colombian region.

---

[7]Correlation coefficients of highway growth and the number of firms and highway growth and the average size of firms are 0.27 and 0.30 respectively across the sample.

IV Econometric Methodology

IV.I Unobserved Endogeneity Bias in Inputs and the GMM Methodology

The estimation of production functions of firms may be affected by biases if endogeneity issues arise, and these are not controlled for. In the context of this study endogeneity issues can arise if (1) highways are extended particularly in regions where high output growth is expected, (2) there are omitted variables that simultaneously influence both the independent input variables and the value-added output variable, and additionally when (3) there are any expected temporary shocks to the firm's productivity that translate into changes in input choices. For the empirical analyses of this paper, we acknowledge that all of these issues may influence the results and hence have to be econometrically addressed. Further, all of these possible sources of biases have been recognized in the literature researching firm performance and public investments. The most established methods to address these issues have been the construction of instrumental variables that rely on historical data[8] or alternatively econometric methods that use the advantages of dynamic panel data that allow to follow firms over time to design a set of internal instrumental variables[9].

In the context of dynamic panel data, the pooled OLS estimator delivers biased results as it does not control for unobserved heterogeneity, endogenous variables and the dynamic autocorrelation of the error. This estimator is therefore unsuitable for the analysis of our data. Compared to the pooled OLS model, the fixed effects (FE) model allows to control for any unobserved heterogeneity across the observations by differencing out any unobserved time invariant factors. Similar to the pooled OLS estimator, the FE model does also not control for endogenous variables, a possible autocorrelation of the errors, or the high level of persistence of the independent variables. Thus, the application of the FE estimator to dynamic panel data remains problematic, especially for data with a large number of observations across a relatively small amount of time periods. The demeaning process employed by the FE estimator creates a correlation between the regressor and the error by subtracting the individual's mean of the dependent and independent variables from their respective variables so that the estimated coefficients are expected to be downwards biased. This is generally referred to as the Nickell bias (Nickell, 1981).

The Difference GMM estimator first differences the original equation to be estimated, and hence removes any unobserved time invariant factors that would otherwise result in a bias from an omitted variable. Subsequently, the estimator instruments the first differences with the lagged values of the

---

[8]See for example Duranton (2015) and Holl (2012)
[9]See Arellano and Bond (1991)

endogenous regressors. However, analyses of firm panel data have identified that input variables are often persistent over time in firm production [10]. Thus, lagged levels are only weak instruments for the first differences in the regressions. The system GMM [11] specification therefore adds the additional assumption of zero correlation between the fixed effects and the differences of the explanatory variables. This method employs both lagged values of the explanatory variables to instrument for current differences, and lagged differences as instruments for current levels for all variables hypothesised to be endogenously determined. The system GMM method offers the additional advantage that it performs better for data with a large number of observations and a finite time horizon; hence it is the preferable GMM estimator for our data consisting of 4,023 observations over 10 years. Further advantages are, given that there exists no correlation across the individual units, it allows to control for heteroscedasticity and autocorrelation within units and their errors. Additionally, in the absence of historical instrument variables, the construction of internal instruments of the GMM methodology allows to control for endogeneity issues in the right-hand side variables, and hence prevents a bias stemming from this source. In the context of our analysis, this allows to control for any unobserved shocks that influence the input choices of the firms.

Throughout the analyses discussed in Section 5, we employ multiple estimation techniques for the results of the whole sample but rely mostly on System GMM results for the subsample analyses; FE estimation results are used as lower-bound benchmarks for System GMM results. The above discussion, and the conclusion drawn from the Monte Carlo simulation of Section 4.3, lead to the conclusion that results obtained with the use of the Difference GMM estimator should only be cautiously interpreted for the work with pseudo-panel data. For this reason, we only provide results obtained with this method in the tables outlining estimation results for the whole sample.

IV.II   Estimation Strategy

We assume that firm output is a function of the standard input factors, capital and labour, and the additional input factors of energy, raw materials and road transportation infrastructure. The underlying hypothesis is that improvements in transportation infrastructure directly reduce input factor costs for firms and hence result in output growth and increased firm level TFP. Furthermore, reductions in transport costs lower the distribution costs for final products and therefore increase the amount of economic mass the firm can access ("effective density"). Additional effects arise through increases in industry level competition resulting in further industry wide TFP improvements.

---

[10]See Blundell and Bond (2000) for further details
[11]See Arellano and Bover (1995) and Blundell and Bond (2000)

The estimation strategy of the firm's output is an extension to the standard neoclassical Cobb-Douglas production function and is represented by

$$Y_{it}(K, L, E, M, , H) = K_{it}^{\beta_{K_t}} L_{it}^{\beta_{L_t}} E_{it}^{\beta_{E_t}} M_{it}^{\beta_{M_t}} H_{rt}^{\beta_{H_t}} e^{\varepsilon_{it}} \tag{1}$$

with

$$\varepsilon_{it} = \rho\varepsilon_{i,t-1} + \mu_i + \tau_t + \epsilon_{it} \tag{2}$$

where Y is the deflated gross value of the output, K is the capital stock, L is the number of permanently employed staff, and E and M are energy and raw materials used respectively in the firm's production. H represents the highway stock for firm i at time t in region r. μ_i represents a firm-specific unobservable time-invariant productivity term, and τ_t captures any unobservable shocks affecting all firms in a given year. The composite error term is further composed of an autocorrelated term $\rho\varepsilon_{i,t-1}$ and the true error ε_{it}.

A transformation of (1) to its logarithmic form yields

$$ln\,Y_{it}(K, L, E, M, H) = \beta_{K_t}ln\,K_{it} + \beta_{L_t}ln\,L_{it} + \beta_{E_t}ln\,E_{it} + \beta_{M_t}ln\,M_{it} + \beta_{H_t}ln\,H_{rt} + \varepsilon_{it} \tag{3}$$

iterating (3) back by a period and solving for $\varepsilon_{i,t-1}$ results in

$$\varepsilon_{i,t-1} = ln\,Y_{it-1} - (\beta_{K_{t-1}}ln\,K_{i,t-1} + \beta_{L_{t-1}}ln\,L_{i,t-1} + \beta_{E_{t-1}}ln\,E_{i,t-1} + \beta_{M_{t-1}}ln\,M_{i,t-1} + \beta_{H_{t-1}}ln\,H_{r,t-1})$$
$$\tag{4}$$

Substituting (4) into (2) and explicitly including all components of the error term transforms (3) into an ARDL model of the first order:

$$ln\,Y_{it}(K, L, , E, M, H) = \rho ln\,Y_{i,t-1} + \beta_{K_t}ln\,K_{it} + \alpha_{K_{t-1}}ln\,K_{i,t-1} + \beta_{L_t}ln\,L_{it} + \alpha_{L_{t-1}}ln\,L_{i,t-1}$$
$$+ \beta_{E_t}ln\,E_{it} + \alpha_{E_{t-1}}ln\,E_{i,t-1} + \beta_{M_t}ln\,M_{it} + \alpha_{M_{t-1}}ln\,M_{i,t-1} + \beta_{H_t}ln\,H_{rt}$$
$$+ \alpha_{H_{t-1}}ln\,H_{r,t-1} + \mu_i + \tau_t + \epsilon_{it} \tag{5}$$

where $\qquad \alpha_{I_{t-1}} = -\rho\beta_{I_{t-1}} \qquad with\,I = K, L, E, M, H\,and\,t = 1,\dots,T$

This first-order autoregressive distributed lag ARDL(1) model specification allows for dynamic effects that arise when adjustments of the firms' output and input choices to changes in the highway

infrastructure are not contemporaneous[12].

The aggregated firm data used for this paper originates from an annually repeated cross-sectional survey. It was provided aggregated at the three-digit ISIC code within each region, therefore the data consisted of one annual observation for each industry within each region. As the underlying data stems from a repeated cross-section, it cannot be assumed that the participating firms remain identical, and thus their numbers constant over time. In order to estimate firm level effects, we follow the pseudo- panel methodology first developed by Deaton (1985) and introduce it as a possible solution for the estimation of firm production functions in the absence of true firm level data. This method restructures the data so that it allows to follow cohorts consistently over time. Deaton initially developed this method for individual level data to estimate models of consumer demand. Cross-sectional data used for pseudo-panels is required to include information on one, or more, observable and time- invariant variables by which the observations can be grouped into cohorts. Subsequently, cohort means for all included variables are constructed, and tracked over time so that the matrix of cohort means forms a panel. This panel of cohort means is referred to as the pseudo-panel. While undoubtedly the major advantage of true panel data is that it enables the identification of precise individual information it can crucially be affected by attrition. Pseudo-panel data, which is constructed from the renewed samples of each year, does not suffer from this issue.

We use the three-digit ISIC code, the region identifier, the year and the information on the number of firms to identify the cohorts and to generate mean variables. Equation (5) becomes:

$$
\begin{aligned}
ln\,Y_{ct}(K,\,L,\,E,\,M,\,H) = {} & \rho ln\,Y_{c,t-1} + \beta_{K_t} ln\,K_{ct} + \alpha_{K_{t-1}} ln\,K_{c,t-1} + \beta_{L_t} ln\,L_{ct} + \alpha_{L_{t-1}} ln\,L_{c,t-1} \\
& + \beta_{E_t} ln\,E_{ct} + \alpha_{E_{t-1}} ln\,E_{c,t-1} + \beta_{M_t} ln\,M_{ct} + \alpha_{M_{t-1}} ln\,M_{c,t-1} + \beta_{H_t} ln\,H_{rt} \\
& + \alpha_{H_{t-1}} ln\,H_{r,t-1} + \mu_c + \tau_t + \epsilon_{ct}
\end{aligned}
\tag{6}
$$

with

$$
I_{ct}^{\alpha_{I_t}} = \bar{I}_{it}^{\alpha_{I_t}} \qquad with \quad I = Y,\,K,\,L,\,E\,M \quad and \quad t = 1,\ldots,T
\tag{7}
$$

where c represents an industry-region cohort, t represents the year, and r denotes the region. Assuming that the size of the cohorts is sufficiently large and the composition relatively stable across the

---

[12]For an additional discussion of the use of ARDL models in the context of roads see Jiwattanakulpaisarn *et al.* (2012)

years, the yearly cohort average of the firm-specific time-invariant effects can be transformed into an industry-region specific unobserved time-invariant effect μ_c which allows to control for unobserved heterogeneity between the cohorts.

If the data exhibit a relatively large degree of within-cohort variation compared to the across-cohort variation, the resulting pseudo-panel estimates may be less efficient than those of the underlying true panel. If the degree of within-cohort variation is relatively small however the loss of efficiency is small. We include cohort-specific effects into our analysis to control for any unobserved between-group heterogeneity across observations. The remaining unobserved between-group heterogeneity is not assumed to be substantial.

Each observation in the subsequent analysis is thus the mean firm of an industry-region cohort at time t and hence allows us to estimate the average effect of road infrastructure on firm output.

IV.III   A Monte Carlo Experiment

In order to assess the validity of estimates based on pseudo-panels in the context of firm data, we conduct a Monte Carlo experiment to compare the differences in the performance of estimators based on a true panel compared to those based on a pseudo-panel constructed from the underlying true panel. While there exists a small body of literature assessing the validity of pseudo-panel estimations in the context of individual or household data[13], the performance of pseudo-panel estimations has not been investigated for firm data and production function estimations. For the estimation of production functions, the literature has highlighted unobserved heterogeneity across firms as one of the main sources of endogeneity affecting the estimation results. This issue is further exacerbated if the unobserved heterogeneity affects the choice, or level, of input variables. Additionally, a possible autocorrelation in the error may introduce a further bias into the estimation of production functions. The established standard panel data methods have shown to control for these factors with different degrees of success, the motivation for our Monte Carlo simulation is hence to assess the performance of these methods in the context of pseudo-panel data to investigate whether these methods are also suitable to control for the above-mentioned issues if the data is not true panel data. An additional motivation for this simulation is the assessment of the performance of logarithmic variables generated from averaged variables as employed by the pseudo-panel used for the empirical estimations of this paper.

The model set up follows a Cobb-Douglas ARDL(1) two input production function structure:

---

[13]See for example Devereux (2007)

$$Y_{it}(X_1, X_2) = X_{1it}^{\beta_{X_{1it}}} X_{2it}^{\beta_{X_{2it}}} e^{\nu_i + \omega_{it}} \qquad t = 1, \ldots, T \tag{8}$$

which can equivalently be expressed in its logarithmic form

$$lnY_{it}(X_1, X_2) = \beta_{X_{1it}} lnX_{1it} + \beta_{X_{2it}} lnX_{2it} + \nu_i + \omega_{it} \tag{9}$$

The variables X_1 and X_2 present the input factors for firm i at time t. We introduce three possible bias sources into the model: a relationship between the lagged dependent variables and the current independent variable, a term capturing unobserved but time-invariant heterogeneity, which affects the dependent variable both directly and indirectly through its influence on the independent variable, and serial correlation in the error. We model the development of both the independent variables in their logarithmic form according to (10). Serial correlation in the composite error term is introduced by including an autocorrelated shock ω_{it} which is independent but exhibits the same variance across the sample; this is expressed in the variables' logarithmic form as described in (11). The parameter ν_i represents the unobserved time-invariant effect which is positively correlated with both regressors. ν_i corresponds to a constant productivity term that acts as a shifter within the production function. This parameter influences both the development of the dependent variable in (8) and (9), and that of the independent variable through (10).

$$lnX_{dit} = \alpha lnX_{di,t-1} + \gamma lnY_{i,t-1} + \delta\nu_i + \varepsilon_{it} \qquad d = 1, 2 \tag{10}$$

$$\omega_{it} = \rho\omega_{i,t-1} + \epsilon_{it} \tag{11}$$

with

$$\varepsilon_{it} \sim N(0,1) \quad \epsilon_{it} \sim N(0,1) \quad \nu_i \sim N(0,1) \quad \omega_{i1} \sim N(0,1)$$

Endogeneity frequently occurs in empirical data in the context of production functions, and often leads to biases and inconsistencies of the estimates generated across different estimators. We investigate the magnitude of this issue by generating endogenous explanatory variables with $\alpha > 0$ and $\gamma > 0$ according to (9).

The model follows the classical ARDL(1) structure where the coefficient of the lagged dependent variable is determined by the level of autocorrelation within the composite error ω_{it}; the coefficients

of the lagged independent variables are determined by both the coefficient of the independent variable at time $t-1$ and the level of autocorrelation in the error. The two explanatory variables are generated with relative differences in the parameters α, γ and δ.

We generate 810 firms across 10 time periods over 1000 Monte Carlo trials. The size of the 108 constructed pseudo-panel cohorts is uniformly distributed across the range of 5 to 10 observations. The model's parameters are chosen to present the level of autocorrelation observed in true firm data with autocorrelation levels of 0.8 and 0.9 for the exogenous variables 1 and 2 respectively. The parameters β_{X_1} and β_{X_2} are set at 0.9 and 0.6 respectively. We further set the autocorrelation within the error term to 0.6. We generate a panel with a length of 20 observations for each unit, and subsequently ignore the first 10 observations for the calculation.

The coefficients are estimated with four standard panel methods: Pooled OLS, the Fixed Effects estimator, and the Difference and System GMM estimators. Results are listed in Table 2. The results of the Pooled OLS estimator exhibit an upwards bias at the second decimal point in the estimated coefficients of the contemporaneous variables, but only show a negligible bias in the lagged variables' coefficients for both true and pseudo-panel estimates. For the estimates of the pseudo-panel, we further observe a relatively slightly lower bias in contemporaneous variables, and slightly larger bias in the estimates of the lagged coefficients. Standard deviations, and thus the root mean square errors, are reported to be larger for pseudo-panel estimates. The Fixed Effects estimator significantly underestimates the coefficient of the lagged dependent variable. This subsequently results in a bias of the coefficients of the lagged independent variables. These biases have already been documented for true panel methods (Nickell, 1981) and our results do not indicate that these biases systematically differ for pseudo-panel estimations.

However, the results of the Difference and System GMM estimations require a more detailed discussion. For the true panel estimates, both Difference and System GMM only indicate biases for true panel data at the second decimal point, Difference GMM results generally indicate relatively small biases for the estimates of contemporaneous variables, but reveal downwards biases for the coefficients of the lagged dependent and independent variables. The shortcomings of the Difference GMM estimation method, discussed in Section 4.1, are exacerbated when pseudo- panel data is used. If the underlying data is highly persistent over time, then the lagged levels as instruments for the first differences in the regression are only weak instruments. A pseudo-panel employs cohort averages as observations, resulting in a data structure which is evidently more persistent over time than the observations of the underlying true panel. This therefore weakens the link between the lagged

Table 2: Monte Carlo Simulation for True and Pseudo-Panel Data

		Pooled OLS			Within			Difference GMM			System GMM		
Coefficient:	β_{true}	$\overline{\beta}$	σ	RMSE	$\overline{\beta}$	σ	RMSE	$\overline{\beta}$	σ	RMSE	$\overline{\beta}$	σ	RMSE
					True Panel Estimation								
$\beta_{Y_{t-1}}$	0.6	0.620	0.013	0.024	0.421	0.016	0.180	0.564	0.019	0.041	0.606	0.016	0.017
$\beta_{X_{1t}}$	0.9	0.923	0.016	0.028	0.898	0.018	0.018	0.895	0.023	0.024	0.927	0.015	0.031
$\beta_{X_{1,t-1}}$	-0.54	-0.544	0.020	0.020	-0.381	0.023	0.161	-0.508	0.024	0.040	-0.532	0.021	0.022
$\beta_{X_{2t}}$	0.6	0.628	0.016	0.032	0.598	0.017	0.017	0.598	0.019	0.019	0.633	0.015	0.036
$\beta_{X_{2,t-1}}$	-0.36	-0.372	0.017	0.021	-0.256	0.018	0.106	-0.337	0.020	0.030	-0.368	0.017	0.019
					Pseudo-Panel Estimation								
$\beta_{Y_{t-1}}$	0.6	0.637	0.027	0.046	0.417	0.032	0.186	0.485	0.036	0.121	0.588	0.031	0.033
$\beta_{X_{1t}}$	0.9	0.922	0.036	0.042	0.899	0.040	0.040	0.890	0.046	0.047	0.930	0.034	0.045
$\beta_{X_{1,t-1}}$	-0.54	-0.549	0.043	0.044	-0.365	0.050	0.182	-0.425	0.051	0.126	-0.506	0.045	0.056
$\beta_{X_{2t}}$	0.6	0.604	0.034	0.034	0.569	0.037	0.048	0.561	0.041	0.057	0.614	0.032	0.034
$\beta_{X_{2,t-1}}$	-0.36	-0.353	0.036	0.037	-0.241	0.039	0.125	-0.273	0.039	0.095	-0.330	0.036	0.047

levels and first differences even further for pseudo-panel data. For pseudo-panels, which have been generated from a true panel data that already exhibits large persistence in the independent variables, this will be particularly noticeable. This increased persistence hence worsens the performance of the Difference GMM estimator for pseudo-panel data.

For true panel estimates, System GMM results, similar to the POLS estimates, exhibit an upwards and downwards bias at the second decimal for the estimates of the contemporaneous variables and lagged coefficients respectively. Furthermore, the System GMM estimate of the coefficient of the lagged dependent variable exhibits the lowest bias across all estimators. While this coefficient reveals an upwards bias for true panel data, it exhibits a downwards bias under pseudo-panel data, however the magnitude of these biases remains very small, and thus negligible.

Across the different estimators, the results reveal a particular degree of heterogeneity across the estimates for the coefficient of the lagged dependent variable, where the System GMM estimates reveal the lowest bias for both true and pseudo-panel estimations. Furthermore, while estimates of the contemporaneous variables generally reveal only negligible biases across all estimators, there exists a relatively larger degree of heterogeneity for the estimates of the lagged dependent variables. An overall comparison of the results of the true and pseudo-panel estimations indicate some, but not significant differences, however there are some deviations at the second and third decimal point; we deem these to be within an acceptable range. The exception is the performance of the Difference GMM estimator which exhibits noticeably larger biases for pseudo-panel data. Further, the pseudo-panel results indicate a loss in efficiency which is reflected in a generally larger standard deviations

and root mean square errors. These results allow us to conclude that the results based on pseudo-panels do not suffer from a crucial bias and can hence be interpreted as valid in the context of production functions.

V   Results

V.I   Baseline Results

Table 3 reports the findings of the estimations of the static and dynamic production function specifications. Column (1) reports the results for the static OLS production function estimation. All input factors except the highway stock, which is negative and insignificant, have the expected sign and are highly significant in this specification. It is reasonable to assume that firms require time to adjust to changes in the transportation infrastructure, i.e. extensions to the existing highway network, hence the results from the dynamic production function model are provided in column (2)[14]. The results show that the coefficients of all the input factors, except highways, have the expected sign and are all highly significant. The coefficients of highways are negative for the contemporaneous and positive for its lagged version, however insignificant for both. Additionally, firms' output appears to be highly autocorrelated. As the results from the pooled OLS (POLS) estimation may include an upwards bias due to the possible endogeneity issues discussed in Section 4.1, the model is tested additionally with the fixed effects model and two different GMM specifications.

The results of the Fixed Effects (FE) estimation model are presented in column (3). Almost all FE coefficients are smaller than those reported under pooled OLS, this particularly affects the coefficients of the lagged variables. As outlined in Section 4.1, the FE estimates are expected to suffer from a downward bias in this context. The results reveal that current transportation infrastructure is insignificant, whereas the coefficient on the lagged highway stock is positive and significant at the 1 per cent level. This provides evidence for the hypothesis that firms' adjustment processes to transportation infrastructure expansions require time. Our results indicate that a 10 per cent increase in the highway stock of a region results in a private sector output growth of manufacturing firms of 1.50 per cent in the subsequent period. The differences in the underlying sample sizes for the calculations of columns (1) in comparison to those of (2) and (3) result predominantly from the requirement of an available lag for the dynamic regressions of the dynamic of (2) and (3). Consequently, both of these estimators will drop the first year from the sample and start the

---

[14]Models including different lag lengths have been tested, subsequently AIC and BIC have been used to identify the optimal lag length with one lag

regression with the variable values from year two with their respective lags from year 1.

To correct for the possible biases in the POLS and FE models, the Difference and System GMM estimators are additionally employed in columns (4) and (5) of Table 3 respectively. Another crucial advantage of employing the GMM methodology results from its construction of internal instruments. In our model, this allows highways to be treated as an additional input factor whose possible endogeneity is treated by the internal GMM instruments. The Difference GMM specification reports the coefficients for all input variables with the expected sign. Highway infrastructure, which is only significant in its lagged value, is slightly lower but in line with the results reported under FE. The results from the System GMM specification are reported in column (5). The reported coefficients of capital, labour, energy and materials are all positive in current levels, negative in lagged values and similar in magnitude than those reported under Difference GMM. The estimated coefficient of highway infrastructure is in line to the coefficient estimated with Difference GMM and FE. Current transportation infrastructure appears to be insignificant for the production, while the lagged level indicates a positive and highly significant relationship. A 10 per cent increase in transportation infrastructure would result in an output growth of 1.53 per cent in the manufacturing sector in the following year. The second point to note is the magnitude of this effect. The mean of the reported output elasticities of transportation infrastructure in the context of developed countries is reported to be around 0.06[15], less than half of the reported coefficient of our analysis. Therefore, the results here provide support for the hypothesis that output elasticities of transportation infrastructure can be substantially higher for developing and emerging economies.

The employed sample sizes for the GMM estimators are affected by the missing data points within unbalanced panels, i.e. in cases where a variable is missing for one period, its differences cannot be calculated for two periods, which reduces the suitable sample size. This particularly affects the Difference GMM estimator, for which the sample size is reported for the transformed data set. The estimations based on the System GMM estimator report the sample size of the underlying untransformed sample, and therefore are not affected by this issue.

Our results indicate a noticeable similarity between the results of Fixed Effects, Difference GMM and System GMM. This could be indicative of the existence of only weak endogeneity, which may not be substantial enough to cause a significant bias. Alternatively, this could be attributed to ineffective internal instruments employed by GMM.

However, contrary to expectations, the System GMM does not yield much higher estimates for

---

[15]See Melo *et al.* (2013) for a comprehensive review of the literature

Table 3: Empirical Results from Static and Dynamic Production Functions

	POLS	POLS	Fixed Effects	Differene GMM	System GMM
Dependent Variable: Ln(Output)$_t$	(1)	(2)	(3)	(4)	(5)
Ln(Output)$_{t-1}$	-	0.881***(0.016)	0.486***(0.037)	0.512***(0.056)	0.536***(0.054)
Ln(Capital)$_t$	0.233***(0.021)	0.093***(0.019)	0.093***(0.018)	0.098***(0.028)	0.092***(0.018)
Ln(Capital)$_{t-1}$	-	-0.065***(0.018)	-0.047***(0.016)	-0.039**(0.019)	-0.051***(0.017)
Ln(Labour)$_t$	0.074***(0.021)	0.049***(0.015)	0.040**(0.016)	0.045*(0.025)	0.040**(0.016)
Ln(Labour)$_{t-1}$	-	-0.043***(0.016)	-0.029*(0.017)	-0.032**(0.015)	-0.032*(0.018)
Ln(Energy)$_t$	0.090***(0.019)	0.160***(0.020)	0.180***(0.026)	0.203***(0.049)	0.179***(0.026)
Ln(Energy)$_{t-1}$	-	-0.147***(0.020)	-0.082***(0.021)	-0.099***(0.020)	-0.090***(0.022)
Ln(Materials)$_t$	0.583***(0.021)	0.623***(0.021)	0.615***(0.023)	0.662***(0.032)	0.615***(0.023)
Ln(Materials)$_{t-1}$	-	-0.557***(0.023)	-0.306***(0.028)	-0.317***(0.038)	-0.337***(0.037)
Ln(Highways)$_t$	-0.020(0.018)	-0.041(0.044)	0.023(0.060)	0.016(0.084)	0.030(0.061)
Ln(Highways)$_{t-1}$	-	0.040(0.044)	0.150***(0.049)	0.131**(0.064)	0.153***(0.047)
Cohort FE	N	N	Y	Y	Y
Time FE	Y	Y	Y	Y	Y
AR1	-	-	-	0.000	0.000
AR2	-	-	-	0.265	0.278
Hansen	-	-	-	0.139	0.000
Number of Instruments	-	-	-	264	771
Observations	4023	3490	3490	3016	3490
R^2	0.963	0.991	0.855	-	-

(where ***, **, * indicate significance at the 1%, 5% and 10% level respectively. Robust standard errors corrected for clustering at the cohort level are provided in parenthesis)

the lagged dependent variable than the Difference GMM estimator. In order to investigate the effectiveness of GMM instruments, we examine the reduced form regressions for first differences and for levels as in Blundell and Bond (2000). We find that in the reduced form of first differences, which relates the first difference of the variables to its lags, the instruments are jointly significant for all variables, except energy. We would therefore expect that the Differenced GMM estimator performs well for all variables except energy. In the reduced form for the levels regression, which relates the first lags to lagged differences of the variables, the instruments are jointly significant for all variables except capital. Hence, we do not expect the System GMM estimator to perform better than Difference GMM for the capital coefficient and but to perform better for energy. Overall, these regressions do not lead us to conclude that the System GMM should not be employed for our data.

V.II.I   Robustness Test I: Additional Controls

To test for the possibility of regional agglomeration economies driving the results of the transportation infrastructure elasticities, employment density and per capita income are included as additional controls. Agglomeration economies describe the productivity benefits that accrue to

firms located in areas with a higher density of economic activity. Sharing of input factors, labour pooling and knowledge spillovers are all representatives of these productivity enhancing benefits termed as agglomeration economies. Areas that have a higher density of economic activity may also have higher growth in highways if growth of economic productivity is expected there; hence, agglomeration economies, rather than highway stock, may be driving the results. The inclusion of employment density allows to control for this issue. If highways are placed in areas where economic growth is expected, then a positive trend in economic performance rather than changes in the infrastructure may explain the results; the inclusion of per capita income allows to gain insight into this issue. In the context of this study, this inclusion of this variable is of further interest given that the 2002 National Development Plan formulates higher road investments in areas with relatively larger national economic importance measured in GDP[16] (Departamento Nacional de Planeación/DNP, 2002).

Additional time-invariant effects for the presence of a seaport and the number of large cities within a region are added to the regression. The former allows to gain insight into the hypotheses that larger benefits of roads may be accrued in regions with an important port given a possibly larger volume of trade or due to the fact that more productive firms have a preference to be situated near a port to reduce transport times and costs. The latter variable allows to control for the possible importance of the distribution of economic activity and the role of cities in increasing firms' output. If increases in the highway network result in heightened levels of competition, then a possible result would be that only the most successful firms survive in the market. This would reduce the number of firms in the market and increase the overall productivity level and output growth exhibited by firms. Hence, it may be the increased competition following highway network expansions rather than the highway expansion itself that drives the results. Preliminary tests have not revealed any strong evidence to support this hypothesis[17]. We investigate the possibility of this further by adding the number of firms included in one cohort level observations and the average firm size, in terms of its employment pool, as additional controls to the estimations.

Table 4 reports the results under pooled OLS (column (1)), Fixed Effects (column (2)) and Difference and System GMM (columns (3) and (4) respectively). The majority of the estimated elasticities across all methods remain similar to those identified for the baseline model of Table 3. Similar to Table 3, highways are only significant in their lagged and not in their contemporaneous values. The inclusion of additional controls decreases the estimated effect of highways by approxi-

---

[16]See Section 3.1
[17]See Section 3.2

mately 0.02 under the FE and System GMM models. Following the introduction of controls to the estimations, a 10 per cent increase in the highway stock would result in an approximate growth in firm's output of 1.3 per cent for both of these methods. Additionally, while lagged highways remain highly significant under FE and GMM and insignificant for POLS, the estimated effect identified with Difference GMM becomes insignificant once additional controls are introduced to the model.

Table 4: Robustness Test I - Additional Controls

Dependent Variable: Ln(Output)_t	POLS (1)	Fixed Effects (2)	Difference GMM (3)	System GMM (4)
Ln(Output)_{t-1}	$0.876^{***}(0.016)$	$0.485^{***}(0.037)$	$0.485^{***}(0.057)$	$0.521^{***}(0.049)$
Ln(Capital)_t	$0.094^{***}(0.019)$	$0.093^{***}(0.018)$	$0.101^{***}(0.026)$	$0.092^{***}(0.018)$
Ln(Capital)_{t-1}	$-0.063^{***}(0.018)$	$-0.048^{***}(0.016)$	$-0.035^{**}(0.018)$	$-0.051^{***}(0.016)$
Ln(Labour)_t	$0.050^{***}(0.015)$	$0.038^{**}(0.017)$	$0.063^{**}(0.032)$	$0.039^{**}(0.017)$
Ln(Labour)_{t-1}	$-0.044^{**}(0.016)$	$-0.029^{*}(0.018)$	$-0.029^{*}(0.016)$	$-0.031^{*}(0.018)$
Ln(Energy)_t	$0.160^{***}(0.020)$	$0.178^{***}(0.026)$	$0.164^{***}(0.038)$	$0.178^{***}(0.026)$
Ln(Energy)_{t-1}	$-0.147^{***}(0.020)$	$-0.081^{***}(0.021)$	$-0.098^{***}(0.020)$	$-0.087^{***}(0.022)$
Ln(Materials)_t	$0.624^{***}(0.021)$	$0.614^{***}(0.023)$	$0.641^{***}(0.030)$	$0.614^{***}(0.023)$
$\text{Ln(Materials)}_{t-1}$	$-0.554^{***}(0.023)$	$-0.305^{***}(0.028)$	$-0.304^{***}(0.039)$	$-0.327^{***}(0.035)$
Ln(Highways)_t	$-0.042(0.044)$	$0.021(0.060)$	$-0.046(0.084)$	$0.022(0.060)$
$\text{Ln(Highways)}_{t-1}$	$0.025(0.044)$	$0.130^{***}(0.050)$	$0.100(0.070)$	$0.131^{***}(0.049)$
Regional Controls				
$\text{Ln(Employment Density)}_t$	$0.004(0.066)$	$0.008(0.090)$	$0.026(0.106)$	$0.012(0.088)$
$\text{Ln(Employment Density)}_{t-1}$	$-0.003(0.066)$	$-0.072(0.075)$	$-0.077(0.078)$	$-0.065(0.074)$
Ln(GDP/Capita)_t	$0.234^{**}(0.112)$	$0.235^{*}(0.124)$	$0.105(0.161)$	$0.231^{*}(0.122)$
$\text{Ln(GDP/Capita)}_{t-1}$	$-0.235^{**}(0.113)$	$-0.239^{**}(0.117)$	$-0.174(0.125)$	$-0.241^{**}(0.117)$
Seaport FE	Y	Y	Y	Y
Urbanisation FE	Y	Y	Y	Y
Cohort Controls				
Number of Firms	$0.0001^{*}(0.0001)$	$-0.0003(0.0004)$	$-0.0001(0.001)$	$-0.001^{*}(0.0004)$
Average Firm Size	$-0.0001(0.0001)$	$0.0001(0.0002)$	$-0.0002(0.0004)$	$0.00001(0.0002)$
Cohort FE	N	Y	Y	Y
Time FE	Y	Y	Y	Y
AR1	-	-	0.000	0.000
AR2	-	-	0.265	0.278
Hansen	-	-	0.139	0.000
Number of Instruments	-	-	264	771
Observations	3490	3490	3016	3490
R^2	0.991	0.855	-	-

(where ***, **, * indicate significance at the 1%, 5% and 10% level respectively. Robust standard errors corrected for clustering at the cohort level are provided in parenthesis)

The estimated coefficients for employment density are very small and insignificant for both the current and lagged period under both estimation methods. Hence, it can be concluded that output elasticities of the highway stock reported previously are not explained by agglomeration economies. Per capita GDP however is reported as significant across all models, with Difference GMM being the exception for both contemporaneous and lagged values. The positive and significant results for per capita GDP and the reduction of the estimated highway elasticity, might imply that the magnitude of the estimated highway coefficient is in part due to a general trend of economic growth during the period investigated. However, given the AR1 structure of the model, which generally results in estimates with opposing signs for both periods included, and the similarity in magnitude of the estimates here, it can be argued that while there exists a positive contemporaneous effect of per capita GDP growth this effect roughly cancels out after two years.

While the number of firms included in each cohort is significant under POLS and System GMM, the magnitude and significance of this effect is very small and hence does not lead us to conclude that this influenced the results of the estimated highway elasticity. Average firm size is very small in magnitude and insignificant across all estimation models[18].

Overall, apart from per capita GDP, the introduction of additional controls does not appear to influence the results. The introduction of per capita GDP however shows that the general regional economic growth trend may explain parts of the estimated highway elasticity, but given that the estimated results remain in line with the results of Table 3, we do not conclude that the results of Table 3 were crucially affected by agglomeration or cohort effects[19].

V.II.II Robustness Test II: Road Density

In order to account for regionally differing characteristics that could influence the effects of roads heterogeneously across Colombian regions, two different specifications of road density are used as alternative measures of transportation infrastructure. First, we use geographic road density, which measures the amount of highway infrastructure (in kilometres) per 100 square kilometres of surface area. This incorporates explicitly the absolute geographic size of each region. This further allows to test whether larger states with a larger road stock and an economy possibly growing at a higher rate influence the results. As an alternative road density specification, we construct a population weighted road density variable. This variable measures the amount of highway infrastructure (in

---

[18]Developments of the number of firms included and the average output level per firm were additionally investigated, but did not result in any significant results.

[19]Detailed results on the separate effect of each additional control introduced can be obtained on request.
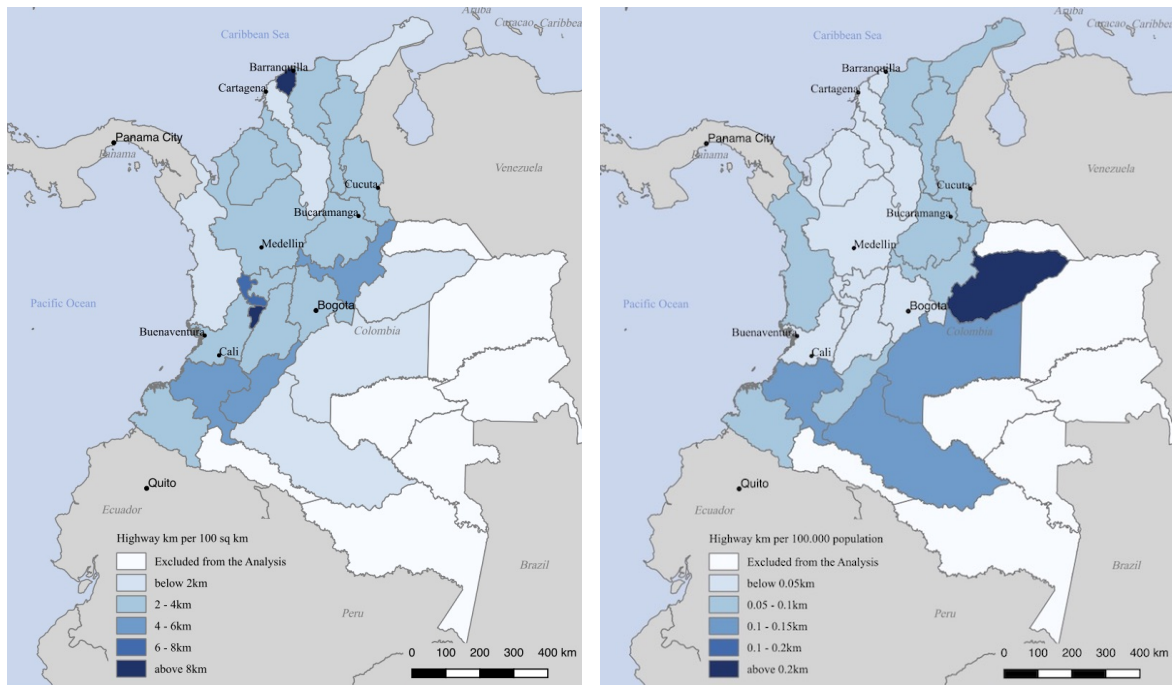
Figure 2: Average Road Density weighted by geography (left) and population (right) (2000 – 2009)
Source: Own elaboration on the basis of the road infrastructure data provided to the authors by INVÌAS; No consistent road data was available for the islands of San Andrés and Providencia, so that these could not be included in this study.

kilometres) per 100,000 population. The inclusion of this variable also tests whether increased congestion effects, which are expected to be present in areas with low road infrastructure relative to the population, influence the previous results. Additionally, as this variable measures road infrastructure relatively, this corrects for the effects of possibly higher road infrastructure investments in regions with higher populations.

Figure 2 provides a graphical representation of both road density variables. The graphs illustrate that while some regions exhibit a geographically dense highway network, predominantly in the centre and towards the coasts, these may not necessarily overlap with those regions that have a dense highway network relative to their population, which are predominantly located towards the East of the country. For those regions, in particular where road density is relatively high for both the geographic and population weighted version, it can be hypothesised that the sole use of the geographic measure might be misleading as in these areas higher levels of congestion should be expected given the large populations. This provides support for the necessity to test the effects of these road density variables separately. Table 5 reports the results under Fixed Effects in columns (1) and (3) and System GMM in columns (2) and (4) for geography and population weighted road

Table 5: Robustness Test II - Road Density

Dependent Variable: Ln(Output)$_t$	weighted by sq. km		weighted by population	
	Fixed Effects	System GMM	Fixed Effect	System GMM
	(1)	(2)	(3)	(4)
Ln(Output)$_{t-1}$	0.485***(0.037)	0.536***(0.048)	0.486***(0.037)	0.530***(0.050)
Ln(Capital)$_t$	0.093***(0.018)	0.092***(0.018)	0.093***(0.018)	0.093***(0.018)
Ln(Capital)$_{t-1}$	-0.047***(0.016)	-0.051***(0.016)	-0.048***(0.016)	-0.052***(0.016)
Ln(Labour)$_t$	0.039**(0.016)	0.039**(0.016)	0.039**(0.016)	0.039**(0.016)
Ln(Labour)$_{t-1}$	-0.029(0.018)	-0.031*(0.016)	-0.029(0.018)	-0.031*(0.018)
Ln(Energy)$_t$	0.179***(0.026)	0.178***(0.026)	0.179***(0.026)	0.178***(0.026)
Ln(Energy)$_{t-1}$	-0.081***(0.021)	-0.090***(0.022)	-0.081***(0.021)	-0.088***(0.022)
Ln(Materials)$_t$	0.615***(0.023)	0.615***(0.023)	0.615***(0.023)	0.615***(0.023)
Ln(Materials)$_{t-1}$	-0.306***(0.028)	-0.337***(0.034)	-0.306***(0.028)	-0.333***(0.035)
Ln(Road Density)$_t$	0.021(0.052)	0.024(0.052)	0.039(0.060)	0.042(0.060)
Ln(Road Density)$_{t-1}$	0.089**(0.042)	0.091**(0.041)	0.156***(0.049)	0.157***(0.047)
Cohort FE	Y	Y	Y	Y
Time FE	Y	Y	Y	Y
Per Capita GDP Growth	Y	Y	Y	Y
AR1	-	0.000	-	0.000
AR2	-	0.357	-	0.316
Hansen	-	0.000	-	0.000
Number of Instruments	-	825	-	825
Observations	3490	3490	3490	3490
R^2	0.855	-	0.855	-

(where ***, **, * indicate significance at the 1%, 5% and 10% level respectively. Robust standard errors corrected for clustering at the cohort level are provided in parenthesis)

density respectively[20].

The estimated coefficients on all inputs factors, except the highway infrastructure variables, remain very similar in magnitude to those of Tables 3 and 4. This allows for the conclusion that the choice of road infrastructure variable does not alter these estimates. Further, while the road density measure differs from columns (1) and (2) to columns (3) and (4) the results show consistently that road infrastructure density affects output growth positively and significantly, albeit only with a time lag. An increase in geographic road density of 10 per cent will lead to a 0.89 to 0.91 per cent growth in firms' output in the following year (columns 1 and 2), and an increase in population based road

---

[20]Given the difference in construction between the road transportation variables employed in this section and the previous, the interpretation of the results of Table 5 do not allow for a direct comparison of these results to those of Tables 3 and 4. Additionally, the population based road density variables comprises two dynamic components, hence an increase in this road density variable implies a relatively higher growth in the road network to population growth.

density of 10 per cent will lead to a 1.56 to 1.57 per cent (columns 3 and 4) rise under FE and GMM estimations respectively.

V.III   Industry Specific Results

The manufacturing sector encompasses a diverse range of manufacturing industries that differ greatly in capital and land - use intensity, the amounts of raw materials and electricity consumed in the production process and the type of final products produced. To test the conjecture that road investments may have a heterogeneous effect on the different manufacturing sectors due to their production differences, our sample was categorised into heavy and light industries. Heavy industries are characterized by capital and land-use intensive production processes, whose final products are often intermediate inputs for other firms, while light industries typically require only limited investment, employ less raw materials and energy and produce goods that are typically final consumer products. The set of light industries for this analysis consists of manufacturing firms of foods and beverages, textiles, fur and wearing apparel, luggage and leather products, wood and cork products, and furniture. The set of manufacturing firms classified as heavy industries for the analysis includes manufacturing of paper and paper products, publishing, printing and media reproduction, production of coke, refined petroleum, nuclear fuel, chemicals, plastic, metal and non-metallic mineral products, and basic metals. Furthermore, included in the heavy industry subsample are the firms engaged in the production of machinery, equipment, motor vehicles, electric apparatus, radio, TV, communication and transport equipment, and the production of medical instruments.

The industry specific estimation results are presented in Table 6. Throughout both estimation methods used, current highway infrastructure remains insignificant and is hence in line with the previous results for both groups of industries. For light industries, the reported effect of lagged road infrastructure stock has roughly the same magnitude across Fixed Effects and System GMM and is significant for both specifications. The results indicate that a 10 per cent increase in highways will result in an output growth in the light industries of 0.62 to 0.63 per cent in the subsequent year under System GMM and Fixed Effects estimations respectively. For heavy industries, the effect of lagged road infrastructure is similar in magnitude and remains highly significant across the estimation techniques, but increases substantially in magnitude compared to the results for light industries and for the whole sample provided in Table 3. Our results suggest that a road expansion of 10 per cent would result in an output growth of heavy industries of 3.51 to 3.58 per cent in the subsequent period (columns (4) and (3) respectively). It is noteworthy to state that while our results

are in line with the previous results in revealing an existing time lag with which road infrastructure expansions affect output growth, the estimated elasticities for heavy industries are more than twice as large as those calculated for the whole sample of manufacturing firms. From these results, we conclude that the benefits from road expansion in Colombia are substantially more accrued to the heavy industries.

Our findings can be compared to the elasticities for trade with respect to intercity highway stock identified by Duranton et al. (2014) for the US. Their findings reveal that a 10 per cent increase in the intercity highway stock raises exports by 5 per cent in weight, while it only has a small and weak effect for exports in value. The authors conclude that roads are an important complement to the production of heavy goods. Repeating this analysis using Colombian trade data, Duranton (2015)

Table 6: Results for Heavy and Light Industries

	Light Industries		Heavy Industries	
	Fixed Effects	System GMM	Fixed Effects	System GMM
Dependent Variable: Ln(Output)$_t$	(1)	(2)	(3)	(4)
Ln(Output)$_{t-1}$	0.439***(0.060)	0.428***(0.066)	0.519***(0.039)	0.545***(0.050)
Ln(Capital)$_t$	0.104***(0.027)	0.104***(0.027)	0.082***(0.018)	0.081***(0.018)
Ln(Capital)$_{t-1}$	-0.057**(0.023)	-0.056**(0.023)	-0.032(0.020)	-0.034*(0.020)
Ln(Labour)$_t$	0.039*(0.020)	0.039*(0.020)	0.036(0.024)	0.036(0.024)
Ln(Labour)$_{t-1}$	-0.040*(0.022)	-0.039*(0.022)	-0.006(0.028)	-0.007(0.027)
Ln(Energy)$_t$	0.198***(0.038)	0.198***(0.038)	0.153***(0.033)	0.152***(0.033)
Ln(Energy)$_{t-1}$	-0.079***(0.030)	-0.077***(0.030)	-0.080***(0.028)	-0.083***(0.028)
Ln(Materials)$_t$	0.617***(0.029)	0.617***(0.029)	0.622***(0.034)	0.622***(0.033)
Ln(Materials)$_{t-1}$	-0.276***(0.044)	-0.269***(0.047)	-0.334***(0.035)	-0.350***(0.041)
Ln(Highways)$_t$	-0.013(0.064)	-0.015(0.063)	0.054(0.133)	0.052(0.131)
Ln(Highways)$_{t-1}$	0.063*(0.035)	0.062*(0.035)	0.358***(0.108)	0.351***(0.104)
Cohort FE	Y	Y	Y	Y
Time FE	Y	Y	Y	Y
Per Capita Growth	Y	Y	Y	Y
AR1	-	0.000	-	0.000
AR2	-	0.243	-	0.499
Hansen	-	1.000	-	0.000
Number of Instruments	-	607	-	593
Observations	1831	1831	1652	1652
R^2	0.856	-	0.860	-

(where ***, **, * indicate significance at the 1%, 5% and 10% level respectively. Robust standard errors corrected for clustering at the cohort level are provided in parenthesis)

reports elasticities for the effect of roads on trade of very similar magnitudes in value and in weight. The reported effect on the exports' value is slightly higher than the author's results for the US, however no further support for the hypothesis of larger productivity benefits from transportation or heavy industries is provided in this paper. In contrast to Duranton (2015) and Duranton et al. (2014), who focus their analyses on roads and trade, our study focuses on output growth and roads. Our findings further support the notion that sectors producing heavy goods exhibit a relatively larger sensitivity to transportation infrastructure.

V.IV   Regional Results

The Colombian regions vary greatly in geographic factors; the Andean mountains run from Northeast the Southwest of the country, the region of the Amazon is located in the South and vast savannah regions exist in the East of the country. Different topographies and soil types generally benefit different industries, hence given the geographic variety of the country, it cannot be assumed that the effects of transportation infrastructure are spatially homogeneous. Additionally, the direct and fast access to a seaport may further affect the results. Port cities are often regional or national economic centres with large domestic markets, and a location choice closer to a port allows firms to reduce transport costs and times. Finally, given that investment in road infrastructure has not been uniformly distributed across the regions[21], it cannot be assumed that its effects are geographically homogeneous. To test for the heterogeneity of transportation infrastructure elasticity across regions, we split the sample into three regional categories: the coastal regions situated on the Atlantic or Pacific coasts, the central regions, and the Eastern peripheral regions. The results are listed in Table 7[22].

The estimated coefficients of capital, labour, energy and materials remain similar across the estimation techniques for each sample, but reveal some degree of variation across the samples. Similar to all previously tested model variations, the estimated elasticities of highways remain insignificant in their contemporaneous values across samples and estimations methods chosen. However, the estimated highway coefficients for the lagged values reveal a degree of heterogeneity across the samples, where while the estimates are in line with the results of Tables 3 - 6 and highly significant for firms located in central regions, they are smaller in magnitude and insignificant for those located in coastal regions. These results indicate that a 10 per cent increase in the highway

---

[21]See section 3.1

[22]The results obtained for eastern peripheral regions were based on a significantly smaller sample size than the other two categories. Given that the results were not robust, they are omitted here but can be obtained by request

Table 7: Regional Results

Dependent Variable: Ln(Output)$_t$	Coastal Regions		Central Regions	
	Fixed Effects	System GMM	Fixed Effects	System GMM
	(1)	(2)	(3)	(4)
Ln(Output)$_{t-1}$	0.469***(0.034)	0.467***(0.039)	0.485***(0.062)	0.501***(0.078)
Ln(Capital)$_t$	0.077***(0.021)	0.077***(0.021)	0.112***(0.027)	0.112***(0.027)
Ln(Capital)$_{t-1}$	-0.048*(0.027)	-0.048*(0.027)	-0.046(0.020)	-0.048**(0.021)
Ln(Labour)$_t$	0.015(0.022)	0.015(0.022)	0.051**(0.021)	0.051(0.021)
Ln(Labour)$_{t-1}$	-0.050*(0.026)	-0.050*(0.026)	-0.017(0.022)	-0.017(0.022)
Ln(Energy)$_t$	0.257***(0.041)	0.257***(0.041)	0.131***(0.028)	0.131***(0.028)
Ln(Energy)$_{t-1}$	-0.046(0.028)	-0.045(0.028)	-0.098***(0.027)	-0.100***(0.027)
Ln(Materials)$_t$	0.584***(0.036)	0.584***(0.036)	0.634***(0.026)	0.635***(0.026)
Ln(Materials)$_{t-1}$	-0.309***(0.031)	-0.308***(0.032)	-0.291***(0.043)	-0.301***(0.052)
Ln(Highways)$_t$	0.055(0.101)	0.055(0.100)	0.042(0.085)	0.044(0.085)
Ln(Highways)$_{t-1}$	0.027(0.340)	0.027(0.336)	0.148**(0.062)	0.149**(0.061)
Cohort FE	Y	Y	Y	Y
Time FE	Y	Y	Y	Y
Per Capita Growth	Y	Y	Y	Y
AR1	-	0.000	-	0.000
AR2	-	0.145	-	0.556
Hansen	-	0.000	-	0.000
Number of Instruments	-	578	-	608
Observations	1605	1605	1830	1830
R^2	0.861	-	0.857	-

(where ***, **, * indicate significance at the 1%, 5% and 10% level respectively. Robust standard errors corrected for clustering at the cohort level are provided in parenthesis)

stock of a region, results in an average output growth of 1.48 to 1.49 per cent for firm in the subsequent period. However, these results are exclusively valid for firms located in central regions. For coastal regions, estimated coefficients on road infrastructure are insignificant for both periods, hence no reliable conclusion can be drawn from these results.

As outlined in Section 3.1, both central and coastal regions have been allocated relatively large shares of the overall road investment budget, hence the disparity in results cannot be explained simply by a larger amount of investment. Figure 1 illustrates the regional disparities across highway growth, however a clear tendency for higher highway growth in central regions cannot be found; additional tests of our data rather show on average higher road growth in coastal regions. Moreover, while the advantage of our dataset is that it provides physical infrastructure data, we do not have

information on the regional distribution of investments across maintenance and road construction projects. For this reason, we cannot exclude the possibility that regional differences in the distribution of funds across road maintenance and new construction projects may influence the results.

The results could also be indicative of a larger economic importance of the central part of the country. We control for the region's per capita GDP to test for the possibility that regional GDP disparities drive the results. We additionally conduct tests on subsamples of economically leading and lagging regions, which only provide weak evidence for stronger effects in economically leading regions[23].

The literature on the regional heterogeneity of the effects of transportation infrastructure remains limited. A noteworthy work in this context has been provided by Storeygard (2016). Storeygard investigates the effects of transportation, proxied for by oil price fluctuations, on income growth of major port cities relative to those located more than 500 kilometres away from the sea. He finds a 7 per cent increase in income of port cities during the investigated time period relative to the cities of the control group. He furthermore identifies an elasticity of city income with respect to transportation costs of -0.28.

The results of this section reveal regional disparities of the effects of road infrastructure on firms' output, and hence highlight the importance of investigating this relationship in more spatial detail. While both the coastal and central regions have received relatively large shares of the national budget during the period investigated, we only find significant results for firms located in the central regions.

VI   Conclusion

This paper investigates the relationship between firm performance and transportation infrastructure in Colombia. In comparison to the previous literature researching this relationship, which predominantly focussed on developed countries or on aggregated data, we provide evidence for the effects of road infrastructure on output growth using aggregated firm data in a developing country context. Our results suggest that roads have substantially larger effects on firms' output growth in developing countries than they do in developed countries. We furthermore identify a time lag with which a firm's production reacts to road stock expansions. We find that an increase in the highway stock of 10 per cent results in additional output growth of 1.3 to 1.5 per cent in the subsequent period. Additionally, we find that the effect of roads on output growth is larger in magnitude for

---

[23]Results available upon request

manufacturing firms in heavy industries with an identified elasticity more than double in magnitude of that estimated for the whole sample. Additional tests of the regional heterogeneity of the effect of transportation infrastructure reveal that the benefits of transportation predominantly accrue to firms located in the central regions of the country. Further robustness tests allow us to identify that regional per capita income may partly drive the estimated elasticities. Additionally, they allow to reject the hypotheses that the results may be driven by agglomeration benefits, cohort factors, or the particular variable chosen to measure transportation infrastructure.

Our paper employs the pseudo-panel methodology as a solution to the absence of true firm level panel data, which is often a problem for empirical work on developing countries. The use of the pseudo-panel methodology allows for the investigation of firm level dynamics in the absence of true firm panel data and hence offers a viable option for research on firms in micro data sparse environments. Further tests do not indicate that a large bias in the coefficients is introduced when using pseudo instead of true panels. Hence, our paper makes a further methodological contribution by investigating the validity of pseudo-panels in the context of production function estimation using aggregated firm data.

While our results support the hypothesis that the effects of transportation infrastructure differ with the state of economic development, further research is required to investigate this relationship in more detail and to examine the underlying mechanisms. It is furthermore important to understand if transportation interacts with the sectoral composition of the economy. In the context of developing countries, it may also be of particular interest to research the relationship of infrastructure and industry shifting.

Acknowledgements

References

Agencia Nacional de Infraestructura/ANI (2013). Autopistas para la Prosperidad completa los precalificados para cinco de sus Proyectos. Available at: http://www.ani.gov.co/article/autopistas-para-la-prosperidad-completa-los-precalificados-para-cinco-de-sus-proyectos-6017. Accessed 10 October 2015.

Arellano, M. and S. Bond (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *The Review of Economic Studies 58*(2), 277–297.

Arellano, M. and O. Bover (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics 68*(1), 29–51.

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

Baldwin, R. E. and T. Okubo (2006). Heterogeneous firms, agglomeration and economic geography: spatial selection and sorting. *Journal of Economic Geography 6*(3), 323–346.

Banerjee, A., E. Duflo, and N. Qian (2012). On the Road: Access to Transportation Infrastructure and Economic Growth in China. The National Bureau of Economic Research.

Barnes, G. and P. Langworthy (2003). The Per-mile Costs of Operating Automobiles and Trucks. Report No. MnDot 2003-19, Minneasota Department of Transportation, Minneapolis, MN.

Blundell, R. and S. Bond (2000). GMM Estimations with Persistent Panel Data: An Application to Production Functions. *Econometric Reviews 3*(19), 321–340.

Blyde, J. (2013). Paving the Road to Export: Assessing the Trade Impact of Road Quality. *International Economic Journal 27*(4), 663–681.

Bougheas, S., P. O. Demetriades, and T. P. Mamuneas (2000). Infrastructure, specialization, and economic growth. *Canadian Journal of Economics/Revue Canadienne d'Economique 33*(2), 506–522.

Buendia, V. and Gagan, O. (2012). IJ Infrastructure Investment Guide: Colombia. Infrastructure Journal, London.

Calderón, C. and L. Servén (2004a). The Effects of Infrastructure Development on Growth and Income Distribution. World Bank Policy Research Papers WPS3400, The World Bank, Washington, D.C.

Calderón, C. and L. Servén (2004b). The Output Cost of Latin America's Infrastructure Gap. In W. Easterly and L. Servén (Eds.), *The Limits of Stabilization: Infrastructure, Public Deficits, and Growth in Latin America*, pp. 95 – 118. Stanford University Press and the World Bank.

Calderón, C. and L. Servén (2010). Infrastructure in Latin America. World Bank Policy Research Papers WPS5317, The World Bank, Washington, D.C.

Crafts, N. (2009). Transport infrastructure investment: implications for growth and productivity. *Oxford Review of Economic Policy 25*(3), 327–343.

Deaton, A. (1985). Panel data from time series of cross-sections. *Journal of Econometrics 30*(1-2), 109–126.

Departamento Nacional de Planeación de Colombia/DNP (2002). Plan Nacional de Desarrollo - Hacia un Estado Comunitario (2002 - 2006). Available at: https://colaboracion.dnp.gov.co/CDT/PND/PND.pdf. Accessed 23 October 2017.

Departamento Nacional de Planeación de Colombia/DNP (2006). Plan Nacional de Desarrollo - Desarrollo para Todos (2006 - 2010). Available at: https://colaboracion.dnp.gov.co/CDT/PND/PND_Tomo_1.pdfDepartamento. Accessed 23 October 2017.

Departamento Nacional de Planeación de Colombia/DNP (2010). Plan Nacional de Desarrollo - Prosperidad para Todos (2010 - 2014). Available at: https://www.dnp.gov.co/Plan-Nacional-de-Desarrollo/PND-2010-2014/Paginas/Plan-Nacional-De-2010-2014.aspx. Accessed 23 October 2017.

Devereux, P. J. (2007). Small-sample bias in synthetic cohort models of labor supply. *Journal of Applied Econometrics 22*(4), 839–848.

Duranton, G. (2015). Roads and trade in Colombia. *Economics of Transportation 4*(1-2), 16–36.

Duranton, G. and P. Diego (2004). Micro-Foundations of Urban Agglomeration Economies. In V. Henderson and J. F. Thisse (Eds.), *Handbook of Regional and Urban Economics*, Volume 4, pp. 2063 – 2117. Amsterdam, Nord-Holland.

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

Duranton, G. and Morrow, P.M. and Turner, M. A. (2014). Roads and Trade: Evidence from the US. *The Review of Economic Studies 81*(2), 681–724.

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

Fernald, J. G. (1999). Roads to Prosperity? Assessing the Link Between Public Capital and Productivity. *American Economic Review 89*(3), 619–638.

Ghani, E. and Grover Goswami, A. and Kerr, W. R. (2016). Highway to Success: The Impact of the Golden Quadrilaterial Project for the Location and Performance of Indian Manufacturing. *The Economic Journal 126*(591), 317 – 357.

Gimenez-Nadal, J. and J. A. Molina (2014). Commuting Time and Labour Supply in the Netherlands. A Time Use Study. *Journal of Transport Economics and Policy 48*(3), 409 – 426.

Graham, D. J. (2007). Agglomeration Economies and Transport Investment. Discussion Paper No. 2007/11, OECD/ITF Joint Transport Research Centre Discussion Papers, OECD Publishing, Paris.

Hansen, N. M. (1965). The Structure and Determinants of Local Public Investment Expenditures. *The Review of Economics and Statistics 47*(2), 150 – 162.

Holl, A. (2006). A Review of the Firm-Level Role of Transport Infrastructure with Implications for Transport Project Evaluation. *A Review of the Firm-Level Role of Transport Infrastructure with Implications for Transport Project Evaluation 21*(1), 3–14.

Holl, A. (2012). Market potential and firm-level productivity in Spain. *Journal of Economic Geography 12*(6), 1191–1215.

Jiwattanakulpaisarn, P., R. B. Noland, and D. J. Graham (2012). Marginal Productivity of Expanding Highway Capacity. *Journal of Transport Economics and Policy 46*(3), 333 – 347.

Lahr, M. L., R. Duran, and A. Varughese (2005). Estimating the Impact of Highways on Average Travel Velocities and Market Size. Unpublished, Center for Urban Policy Research, Rutgers University, New Brunswick, N.J.

Lall, S. V. (2007). Infrastructure and regional growth, growth dynamics and policy relevance for India. *The Annals of Regional Science 41*(3), 581–599.

Melitz, M. J. (2003). The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. *Econometrica 71*(6), 1695–1725.

Melo, P. C., D. J. Graham, and R. Brage-Ardao (2013). The productivity of transport infrastructure investment: A meta-analysis of empirical evidence. *Regional Science and Urban Economics 43*(5), 695–706.

Nickell, S. (1981). Biases in Dynamic Models with Fixed Effects. *Econometrica 49*(6), 1417–1426.

Nieto-Parra, S., M. Olivera, and A. Tibocha (2013). The Politics of Transport Infrastructure Policies in Colombia. OECD Development Centre Working Paper No. 316, OECE Paris.

Rodríguez-Pose, A., V. Tselios, D. Winkler, and T. Farole (2013). Geography and the Determinants of Firm Exports in Indonesia. *World Development 44*, 225–240.

Storeygard, A. (2016). Farther on Down the Road: Transport Costs, Trade and Urban Growth in Sub-Saharan Africa. *The Review of Economic Studies 83*(3), 1263–1295.

The World Bank (2004). World Development Report 2005: A Better Investment Climate for Everyone. The World Bank and the Oxford University Press, New York, N.Y.

Tripathi, S. and V. Gautam (2010). Road Transport Infrastructure and Economic Growth in India. *Journal of Infrastructure Development 2*(2), 135–151.

Venables, A. J., J. Laird, and H. G. Overman (2014). Transport Investment and Economic Performance: Implications for Project Appraisal. Department for Transport, London.

Appendix

Additional Descriptive Statistics

The below Table A1 provides additional descriptive statistics of factors employed for this work. The average number of cohorts, firms and employees lists the number of industry-region cohorts, the average number of firms and employees averaged to the regional level across all years employed for the analysis. The average highway growth shows the average growth of the regional highway stock, average geographic road density growth lists the average growth of highway kilometres per 100 sq. kilometres and the average population weighted road density growth measures the average growth of highway kilometres per 100,000 population in each region, all in their natural logarithms, across the years employed for this study.

Table A1: Additional Descriptive Statistics

Departamento	Average Number of Cohorts	Average Number of Firms	Average Number of Employees	Average Highway Growth Growth	Average Geographic Road Density Density Growth	Average Popluation weighted Road	Freq.
Antioquia	49.13	1,460.85	82,169.88	7.33	2.39	0.027	12.20
Atlántico	30.82	360.96	17,191.45	5.48	8.00	0.011	7.66
Bogotá, D.C.	54.63	2,653.63	117,247.6	6.69	3.32	0.009	13.57
Bolívar	16.86	120.37	6,326.60	6.21	1.93	0.027	4.18
Boyacá	7.48	57.62	4,200.42	6.94	4.49	0.083	1.64
Caldas	21.84	172.02	6,403.30	5.73	3.91	0.032	5.39
Caquetá	1.00	4.10	69.4	6.08	0.49	0.10	0.25
Cauca	16.32	103.54	6,072.59	7.21	4.60	0.11	4.03
Cesar	5.55	30.84	1,616.64	6.48	2.91	0.074	1.37
Córdoba	5.63	29.69	1,778.96	6.38	2.36	0.040	1.34
Cundinamarca	27.53	329.67	25,517.26	6.69	3.32	0.009	6.79
Chocó	1.00	3.00	31.50	5.62	0.59	0.063	0.05
Huila	9.22	50.68	816.28	6.74	4.25	0.084	2.29
La Guajira	1.18	3.55	20.55	5.87	1.70	0.054	0.27
Magdalena	8.32	50.30	1,660.90	6.45	2.74	0.055	2.04
Meta	5.44	48.11	1,937.63	6.90	1.16	0.13	1.34
Nariño	8.14	57.36	1,039.94	6.65	2.36	0.050	2.01
Norte de Santander	14.19	146.99	2,550.68	6.76	3.85	0.069	3.50
Quindío	8.08	61.85	621.39	5.11	9.03	0.031	1.99
Risaralda	23.79	182.82	6,832.83	5.71	7.27	0.034	5.89
Santander	30.35	363.19	8.677.85	7.11	4.00	0.063	7.53
Sucre	2.57	14.00	455.61	5.62	2.59	0.036	0.57
Tolima	11.07	120.40	3,204.84	6.33	2.38	0.041	2.68
Valle del Cauca	45.72	1,064.56	48,610.51	6.52	3.05	0.017	11.36
Casanara	2.00	10.00	228.00	6.65	1.73	0.241	0.05