Centre for Globalization Research
School of Business and Management

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

CGR Working Paper 69

Samira Barzin, Sabine D'Costa and Daniel Graham

Abstract

We construct a pseudo-panel of Colombian firms based on the Colombian Annual Manufacturing Survey to study the effects of transportation infrastructure on firm performance in a developing country. Our findings report an output elasticity with respect to road infrastructure of 0.132 to 0.146 across the specifications, which confirms our initial hypothesis that roads are an important driver for private sector output growth. The fact that our results are larger than those reported in the literature for developed countries could suggest that the role of transportation infrastructure is relatively more important for the economy of developing countries. Furthermore, our findings reveal that there exists a time lag with which firms' productions react to changes in the road stock. We interpret these findings as firms requiring time to adjust their production processes to road improvements of at least a year. We furthermore identify that the effect of road infrastructure is particularly large for those manufacturing industries that are capital-intensive and produce heavy goods. Further robustness tests reveal that our results are not driven by the possibility of agglomeration economies or the chosen measurement of transportation infrastructure. We additionally provide Monte Carlo simulations to provide support for the validity of pseudo-panels in the context of firm-level data.

JEL: O18, O14, R42, C15

Keywords: Infrastructure; Roads; Economic Development; Pseudo-Panels; Monte Carlo Simulations; Latin America; Colombia

http://www.busman.qmul.ac.uk/cgr

CGR Working Paper Series

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

Samira Barzin*[+], Sabine D'Costa** and Daniel J. Graham*

*Imperial College London, Department of Civil and Environmental Engineering and the Railway and Transport Strategy Centre
** Westminster Business School, University of Westminster, and SERC

[+] Corresponding author: s.barzin12@imperial.ac.uk

July 2016

Abstract

We construct a pseudo-panel of Colombian firms based on the Colombian Annual Manufacturing Survey to study the effects of transportation infrastructure on firm performance in a developing country. Our findings report an output elasticity with respect to road infrastructure of 0.132 to 0.146 across the specifications, which confirms our initial hypothesis that roads are an important driver for private sector output growth. The fact that our results are larger than those reported in the literature for developed countries could suggest that the role of transportation infrastructure is relatively more important for the economy of developing countries. Furthermore, our findings reveal that there exists a time lag with which firms' productions react to changes in the road stock. We interpret these findings as firms requiring time to adjust their production processes to road improvements of at least a year. We furthermore identify that the effect of road infrastructure is particularly large for those manufacturing industries that are capital-intensive and produce heavy goods. Further robustness tests reveal that our results are not driven by the possibility of agglomeration economies or the chosen measurement of transportation infrastructure. We additionally provide Monte Carlo simulations to provide support for the validity of pseudo-panels in the context of firm-level data.

Keywords: Infrastructure; Roads; Economic Development; Pseudo-Panels; Monte Carlo Simulations; Latin America; Colombia

JEL Codes: O18, O14, R42, C15

1. Introduction

Transportation infrastructure is as a crucial component to economic growth (Tripathi and Gautam (2010) and Crafts (2009)). 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 rich Western and East Asian countries since the 1970s. Additionally, with an average infrastructure spending of 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 kms of highway road infrastructure until the year 2020. The main goal of this project, which is estimated to cost around 25 billion 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 in the northwest of the country. Deemed currently as the most extensive transportation project globally with an expected cost of 7.2 billion 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 this extensive project, we provide an insight into the relationship between road infrastructure and the Colombian economy for the years 2000 to 2009 by conducting an ex-post evaluation. 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

[1] See reports by Agencia Nacional de Infraestructura de Colombia, 2013, Departamento Nacional de Planeacion de Colombia, 2014, and Infrastructure Journal Investment Guide: Colombia, 2012.

cent , results in manufacturing output growth of 1.32 to 1.46 per cent in the subsequent year for the whole sample. Results are significantly larger for heavy industries. These results suggest that firms' production processes require time to adjust to highway expansions. The identified elasticities furthermore indicate that the returns from highway expansions on firms of the private sector 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, discuss the data sources in Section 3, we outline the econometric model for the analysis, including a discussion of the pseudo-panel methodology and the associated Monte Carlo simulations in Section 4 and in Section 5 we analyse the results. Section 6 concludes.

2. An Overview of the Context and 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, 2011). These benefits will have direct positive effects on the level of productivity of firms.

Additional benefits arise due to changes in agglomeration economies and effective density resulting from transportation cost reductions (Graham 2007). These encompass sharing of resources across larger geographical space, more efficient matching between employers and employees, or across business partners, and increased information exchange through knowledge sharing and 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 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 further increase the pressure on surviving firms to increase their productivity level so that the overall degree of average productivity will increase (Baldwin and Okubo, 2006 and Melitz, 2003). This may additionally encourage specialization across

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

firms as a comparative advantage becomes relatively more important to remain in a more competitive market (Bougheas *et al*., 2000). Furthermore, changes in transport costs can also affect the 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 the input factor mix (Holl, 2006). Additionally, if transport cost reductions yield price reductions of the final good, increased demand could be generated (Lahr *et al*., 2005). Moreover, better transportation infrastructure allows for a longer lifespan of the existing capital stock, trucks often depreciate at a lower rate on better quality roads (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 can hence encourage private investments (Crafts, 2009). This will 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 due to 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 criticized heavily on the basis of various econometric issues. As a response more recent papers rely on more complex estimation strategies in order to prevent biases from endogeneity issues such as 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 technique relies on an instrumental variable approach that employs a historical highway plan and a historical railroad map as exogenous factors in determining current highways but not 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 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 bias, she relies on historical data to construct historical instrumental variables for current market access. Her reported estimates of the effect of the growth of market access on output growth range from 0.042 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 because they were geographically located between targeted cities. In order to further prevent any endogeneity bias, he constructs two instrumental variables that are based on hypothetical least costly road paths of the construction of the highway. 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 the economic activity away 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.

Calderón & Servén (2004a) estimate the effects of infrastructure on GDP using a large panel of 120 developed and developing countries from 1960 to 2000. Their infrastructure index includes both infrastructure quantity and quality and their results indicate that GDP growth is positively influenced by all included infrastructure factors. Focusing the analysis on Latin America, Calderón & 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 exceeds those of non-infrastructure capital. They also find that the output gap between the Latin America and East Asia countries throughout the 1980s and 1990s is largely due to 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 affects state-level output growth positively. 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 firms' export propensities of the Indonesian manufacturing sector, Rodríguez-Pose *et al*. (2013) identify the role of transportation infrastructure as a particularly crucial factor determining firms' export patterns. The authors show further that both the road infrastructure of the region where the firm is located and the road infrastructure of neighbouring regions influence firms' export patterns.

While, as outlined above, there exists some research on the effects of transportation infrastructure capital on economic growth for developing countries, to date this literature remains limited. This is particularly the case for firm-level studies. 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 Colombian aggregated firm data and combine this with transportation data to estimate the effects of the road network on firm-level output. This paper 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 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 a pseudo-panel of firm data to focus on the role of roads on output growth.

[3] see for example Hansen (1965)

3. Data

The compiled data set includes data from the capital district of Bogotá and 23 out of the 32 Colombian regions (departamentos). The regions of Casanare, Chocó, Vaupes and the island state of San Andrés y Providencia were excluded due to insufficient economic data. The majority of the regions of Arauca, Amazonas, Guainía, Guaviare, Putumayo are not covered by the national highway network and are hence excluded. Annual data covering the years 2000 to 2009 were used for the analysis.

The main source of data used to measure the road stock is yearly data on the Colombian highway stock (km) per Colombian region provided by the National Roads Institute of Colombia INVÍAS. The information on firms 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 output, capital stock, employment, inventories, raw materials usage, electricity usage, and investments. The data were provided to us aggregated by 3-digit ISIC, Rev. 3 industry for each Colombian region. We employed the information on the number of firms included in each three-digit industrial sector–departamento pair to generate a pseudo panel encompassing 4052 observations[4], where each observation represents an average firm for a given industry in a given region in a given year. 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 these labour market surveys provide information on the working age population, unemployment rates, and the amount of employed among others for each region. The Encuesta Anual Manufacturera, the Gran Encuesta de Hogares and the Encuesta Nacional de Hogares have been obtained through the Departamento Administrativo Nacional de Estadística DANE.

Table 1 provides a summary of the descriptive statistics of the main variables used.
Output, capital and raw materials were provided as measured in thousands of Colombian pesos. In order to compute 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 manufacturing of machinery and equipment, and raw materials are deflated by the annual average manufacturing producer price index. Energy, labour and the highway infrastructure

[4] further details on the pseudo-panel methodology can be found in Section 4

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.

4. Econometric Methodology

4.1 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 hence increases 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.

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

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

with

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

where *Y* is the deflated gross value of the output of a firm, *K* is the capital stock, *L* is the number of permanently employed staff, *E* and *M* are energy and raw materials used respectively, and *H* represents the highway stock for firm i at time t. 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 $\rho\varepsilon_{it-1}$ and the true error ϵ_{it}.

A log-linear transformation of (1) yields

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

iterating (3) back by a period and solving for ε_{it-1} results in

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

Substituting (4) into (2) and explicitly including all components of the error term transforms (3) into an ARDL model of the first order[5]:

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

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[6].

The firm level data underlying this paper stems from an annually repeated cross sectional survey. It was provided aggregated at the three-digit ISIC code within each region, so that the data consisted of one annual observation for each industry within each region. In order to estimate firm level effects, we follow the pseudo-panel methodology first developed by Deaton (1985) that restructured 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. The cross-sectional data is required to include information on one or more observable and time-invariant variables by which the observations are grouped into cohorts. Subsequently cohort means for any variable 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.

[5] note that $\alpha_{It-1} = -\beta_{It}\rho \quad with\ I = K,L,E,M,H\ and\ t = 1,\dots,T$

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

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. Function (5) becomes:

$$\ln Y_{ct}\ (K,L,E,M,H) = \alpha_{Y_{t-1}} \ln Y_{ct-1} + \beta_{K_t} \ln K_{ct} + \alpha_{K_{t-1}} \ln K_{ct-1} + \beta_{L_t} \ln L_{ct} + \alpha_{L_{t-1}} \ln L_{ct-1} + \beta \ln E_{ct} + \alpha_{E_{t-1}} \ln E_{ct-1} + \beta_{M_t} \ln M_{ct} + \alpha_{M_{t-1}} \ln M_{ct-1} + \beta_{H_t} \ln H_{ct} + \alpha_{H_t} \ln H_{ct-1} + \mu_c + \tau_t + \epsilon_{ct} \quad (6)$$

with

$$I_{ct}^{\alpha_{I_t}} = \overline{I_{\iota t}^{\alpha_{I_t}}} \qquad with\ I = K,L,E,M,H\ and\ t = 1,\dots,T \quad (7)$$

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

4.2 A Monte Carlo experiment

In order to assess the validity of estimates based on pseudo-panel data 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 that has been constructed from the underlying true panel. In order to analyse the performance of estimates based on the different versions of the data, we assess the performance of the estimates under different estimators. The model set up follows a Cobb-Douglas ARDL(1) two input production function structure:

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

$$X_{jit} = \alpha X_{ji,t-1} + \gamma Y_{i,t-1} + \delta \nu_i + \varepsilon_{it} \qquad j = 1,2 \tag{9}$$

$$\omega_{it} = \rho \omega_{it-1} + \epsilon_{it} \tag{10}$$

with

$$\varepsilon_{it} \backsim N(0,1) \qquad \epsilon_{it} \backsim N(0,1) \quad \nu_i \backsim N(0,0.3) \qquad \omega_{i1} \backsim N(0,1)$$

The variables X_1 and X_2 present the input factors for firm i at time t. We allow for serial correlation in the composite error term by including an autocorrelated shock ω_{it} which is independent and but has the same variance across the sample. The parameter ν_i represents an unobserved time-invariant effect which is positively correlated with both regressors. ν_i corresponds to a constant productivity term that acts as a shifter of the production function.
Endogeneity is frequently observed in empirical data in the context of production functions, but often results in 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}, and 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, δ and λ.

We consider 400 firms across 10 time periods over 1000 Monte Carlo trials. The pseudo-panel is constructed from 40 cohorts that encompass 10 observations each. The model's parameters are chosen to present the level of autocorrelation observed in true firm data with autocorrelation levels α of 0.95 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 at 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 three standard panel methods: Pooled OLS, the Fixed Effects estimator, and the Difference and System GMM estimators. Results are listed in Table 2. The results indicate that coefficients estimated under Pooled OLS and Fixed Effects over and under estimate the coefficients respectively, whereas the results from the Difference and System GMM do not indicate large biases; comparing the two separate GMM approaches, System GMM generally indicates smaller biases. The results from the true and pseudo panel estimations do not indicate large differences, although there are some differences at the second a third decimal point. Further, the pseudo panel results indicate a loss in efficiency which is reflected in a generally larger root mean square error.

Overall, 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.

4.3 Unobserved endogeneity bias in inputs and the GMM methodology

In the context of estimating production functions of firms, endogeneity issues may arise and bias the estimation results if they 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, if (2) there are omitted variables that simultaneously influence both the independent input variables and the value added output variable, and additionally if (3) any expected temporary shocks to the firm's productivity translate into changes in input choices. 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 use of instrumental variables that rely on historical data[7] or 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[8].

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 and hence it is unsuitable for the analysis of our data. Compared to the pooled OLS model, the 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 or a possible autocorrelation of the errors, or the high level of persistence of the independent

[7] see for example Duranton (2015) and Holl (2012)
[8] see Arellano and Bond (1991)

variables so that the application of the FE estimator to dynamic panel data remains problematic, especially in the context of 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 (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 endogenous regressors, however analyses of firm panel data have identified that input variables are often persistent over time in firm production[9] so lagged levels are only weak instruments for the first differences in the regressions. The system GMM specification[10] adds the additional assumption of zero correlation between the fixed effects and the differences of the explanatory variables. This method employs lagged values of the explanatory variables to instrument for current differences and it uses lagged differences as instruments for current levels. 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 4052 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. In the context of our analysis this allows to control for any unobserved shocks that influence the input choices of the firms.

5. Results

5.1 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 road stock, which is negative, have the expected sign and

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

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)[11]. The results show that the coefficients of all the input factors have the expected sign and are all highly significant. Additionally, firms' output appears to be highly autocorrelated. As the results from the pooled OLS estimation may include an upwards bias due to the possible endogeneity issues discussed in Section 4.3, the model is tested additionally with the fixed effects model and two different specifications of the generalized method of moments (GMM). The results of the Fixed Effects (FE) estimation model are presented in column (3). FE coefficients are smaller than those reported under pooled OLS, this particularly affects the coefficients of the lagged variables. As outlined in section 4.3, 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 1.45 per cent in the subsequent period.

To correct for the possible biases in the POLS and FE models, the Difference (column (4)) and System GMM estimators (column (5)) are additionally employed. The Difference GMM specification reports the coefficients for all input variables with the expected sign. Highway infrastructure, which is reported to be 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 very close to the coefficient estimated with Difference GMM. 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 1.46 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

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

context of developed countries is reported to be around 0.06[12], 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.

Our results indicate a noticeable similarity between the results of Fixed Effects, difference GMM and system GMM. This could be indicative for the existence of only weak endogeneity, which may not be substantial enough to cause a bias. Alternatively, this could be attributed to ineffective internal instruments employed by GMM. The difference GMM estimates are biased towards the fixed effects estimator, which is in line with a finite sample bias in the context of relatively persistent data. However, contrary to expectations, the system GMM does not yield much higher estimates for the lagged dependent variable. 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 expect the system GMM estimator not to perform better than difference GMM for the capital coefficient and 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.

Robustness test I: employment density

To test for the possibility of regional agglomeration economies driving the results of the transportation infrastructure elasticities, employment density is included as a further explanatory variable in the regression. 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 roads if growth of

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

economic productivity is expected there; hence agglomeration economies, rather than highway stock, may be driving the results. Table 4 reports the results for this specification under Fixed Effects in column (1) and under system GMM in column (2). The estimated elasticities under both methods remain similar to those estimated for the model without employment density. The estimated coefficients for employment density are very small and insignificant for both the current and lagged period under both estimation methods. Similar to the model excluding employment density, the estimated coefficient of current highway stock is insignificant while the lagged highway stock is positive and significant. Further, similar to the results reported in Table 3 the estimated coefficient of the lagged highway stock is very similar under Fixed Effects and System GMM.

As employment density remains insignificant under both methods and the reported elasticities for the current and lagged highway stock remain very similar to those reported in the model excluding employment density, it can be concluded that output elasticities of the highway stock reported previously are not explained by agglomeration economies.

Robustness test II: road density

In order to account for the differing sizes of Colombian regions, road density is used as an alternative measure of transportation infrastructure. Road density is a measure of the amount of highway infrastructure (in kilometres) per 100 square kilometres of surface and hence incorporates the absolute size of each region explicitly. This allows to test whether larger states with a larger road stock and an economy possibly growing at a higher rate influence the results. Table 5 reports the results under Fixed Effects in column (1) and under system GMM in column (2). Under FE the results remain very close to those of the original specification in both magnitude and significance. Using road density as an alternative measure of transportation infrastructure also confirms our results of Table 3 that changes in transportation infrastructure affect firms' behaviour only in the subsequent period. While current levels of road density remain insignificant, the lagged levels of road density are positive and significant, where an increase in road density of 10 per cent will lead to a 1.02 per cent increase in firms' output in the following year. Under the system GMM specification the results similarly are of slightly lower but similar size and significance compared to those of table 6. The results indicate that changes in road density only affect firms' productions with a lag, while current levels of road density remain insignificant. These results indicate that an increase in road density of 10 per cent would be associated with an output growth of 1.04 per cent in the subsequent period.

The additional results confirm the previous findings that there exists a positive effect of transportation infrastructure on private sector output and furthermore that there exists a dynamic component to this.

5.2 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 in the type of final products produced. To test the conjecture that road investments may have a different 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 than heavy industries and produce goods that are typically final consumer products. The set of light industries for the 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 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 hence in line with the previous results for both groups of industries. For light industries, the reported effect of lagged road infrastructure stock is very similar across Fixed Effects and system GMM and 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 0.09 per cent in the subsequent year under both specifications. 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 of Table 3. Our results suggest that a road expansion of 10 per cent would result in an output growth of heavy industries of 3.4 per cent in the subsequent period (columns (3) and (4)). 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 increases 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) 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 for heavy industries is provided in this paper. In contrast to Duranton *et al.* (2014) and Duranton (2015) 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.

6. 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 a pseudo-panel of firm data in a developing country context. Our results suggest that roads have larger effects on firms' output growth in developing countries than they do in developed countries with dense transportation networks. 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 output growth of 1.46 per cent in the subsequent period. Additionally, we find that the effect of roads on output growth is larger in magnitude for manufacturing firms in heavy industries with an identified elasticity more than double in magnitude of that estimated for the whole sample. Further robustness tests allow us to reject the hypotheses that the results may be driven by agglomeration benefits or the 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. 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 firm-level 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.

Tables

Table 1
Descriptive Statistics

Variable	Mean	Standard Deviation
Output	14,400	51,600
Capital	10,600	10,600
Employment	49.29	60.18
Energy	2,907	11,400
Materials	8,560	42,900
Highways	794.90	400.44

Table 2

Monte Carlo Simulation for True and Pseudo-Panel Data

Coefficient:	$\boldsymbol{\beta_{true}}$	True Panel Estimation			Pseudo-Panel Estimation		
		$\bar{\beta}$	σ	RMSE	$\bar{\beta}$	σ	RMSE
				Pooled OLS			
$\boldsymbol{\beta_{Y_{t-1}}}$	0.6	0.621	0.015	0.152	0.621	0.045	0.050
$\boldsymbol{\beta_{X_{1t}}}$	0.9	0.896	0.016	0.158	0.897	0.055	0.055
$\boldsymbol{\beta_{X_{1,t-1}}}$	-0.54	-0.552	0.021	0.024	-0.556	0.070	0.071
$\boldsymbol{\beta_{X_{2t}}}$	0.6	0.603	0.013	0.013	0.614	0.044	0.046
$\boldsymbol{\beta_{X_{2,t-1}}}$	-0.36	-0.371	0.014	0.018	-0.380	0.052	0.056
				Fixed Effects			
$\boldsymbol{\beta_{Y_{t-1}}}$	0.6	0.426	0.017	0.175	0.418	0.037	0.186
$\boldsymbol{\beta_{X_{1t}}}$	0.9	0.887	0.017	0.021	0.900	0.062	0.062
$\boldsymbol{\beta_{X_{1,t-1}}}$	-0.54	-0.386	0.022	0.156	-0.380	0.067	0.070
$\boldsymbol{\beta_{X_{2t}}}$	0.6	0.592	0.014	0.016	0.603	0.057	0.057
$\boldsymbol{\beta_{X_{2,t-1}}}$	-0.36	-0.264	0.015	0.097	-0.267	0.047	0.104
				Difference GMM			
$\boldsymbol{\beta_{Y_{t-1}}}$	0.6	0.569	0.023	0.039	0.433	0.039	0.171
$\boldsymbol{\beta_{X_{1t}}}$	0.9	0.895	0.019	0.020	0.900	0.062	0.062
$\boldsymbol{\beta_{X_{1,t-1}}}$	-0.54	-0.507	0.023	0.040	-0.395	0.068	0.160
$\boldsymbol{\beta_{X_{2t}}}$	0.6	0.592	0.026	0.027	0.599	0.052	0.052
$\boldsymbol{\beta_{X_{2,t-1}}}$	-0.36	-0.346	0.015	0.021	-0.274	0.045	0.097
				System GMM			
$\boldsymbol{\beta_{Y_{t-1}}}$	0.6	0.596	0.019	0.019	0.604	0.041	0.041
$\boldsymbol{\beta_{X_{1t}}}$	0.9	0.901	0.015	0.015	0.901	0.055	0.055
$\boldsymbol{\beta_{X_{1,t-1}}}$	-0.54	-0.534	0.019	0.020	-0.544	0.070	0.070
$\boldsymbol{\beta_{X_{2t}}}$	0.6	0.606	0.013	0.014	0.617	0.045	0.048
$\boldsymbol{\beta_{X_{2,t-1}}}$	-0.36	-0.358	0.013	0.013	-0.372	0.051	0.052

Table 3

Empirical Results from Static and Dynamic Production Functions

Dependent Variable: Ln(Output)$_t$	OLS	OLS	Fixed Effects	Difference GMM	System GMM
	(1)	(2)	(3)	(4)	(5)
Ln(Output)$_{t-1}$	-	0.882***(0.019)	0.487**(0.037)	0.517***(0.057)	0.537***(0.054)
Ln(Capital)$_t$	0.234***(0.010)	0.093***(0.016)	0.094***(0.018)	0.097***(0.028)	0.093***(0.018)
Ln(Capital)$_{t-1}$	-	-0.066***(0.016)	-0.048***(0.016)	-0.041**(0.019)	-0.052***(0.016)
Ln(Employment)$_t$	0.074***(0.009)	0.048***(0.014)	0.038**(0.016)	0.041(0.026)	0.038**(0.016)
Ln(Employment)$_{t-1}$	-	-0.042***(0.014)	-0.029*(0.017)	-0.033**(0.015)	-0.031*(0.018)
Ln(Energy)$_t$	0.089***(0.009)	0.160***(0.019)	0.179***(0.028)	0.209***(0.050)	0.179***(0.026)
Ln(Energy)$_{t-1}$	-	-0.147***(0.019)	-0.081***(0.021)	-0.097***(0.020)	-0.090***(0.022)
Ln(Materials)$_t$	0.584***(0.007)	0.624***(0.018)	0.616***(0.023)	0.659***(0.031)	0.616***(0.023)
Ln(Materials)$_{t-1}$	-	-0.558***(0.022)	-0.306***(0.028)	-0.319***(0.038)	-0.337***(0.038)
Ln(Highways)$_t$	-0.019**(0.008)	-0.035(0.044)	0.031(0.059)	0.026(0.085)	0.034(0.059)
Ln(Highways)$_{t-1}$	-	0.035(0.044)	0.145***(0.048)	0.132**(0.067)	0.146***(0.047)
Cohort FE	No	No	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes
AR1	-	-	-	0.000	0.000
AR2	-	-	-	0.272	0.280
Hansen	-	-	-	0.108	0.254
Number of Instruments	-	-	-	264	775
Observations	4050	3512	3512	3033	3512
R^2	0.964	0.992	0.980	-	-

(where ***, **, * indicate significance at the 1%, 5% and 10% level respectively. Robust standard errors in parenthesis)

Table 4

Robustness Test I – Employment Density

Dependent Variable: $Ln(Output)_t$	Employment Density	
	Fixed Effects	System GMM
	(1)	(2)
$Ln(Output)_{t-1}$	0.486***(0.037)	0.536***(0.053)
$Ln(Capital)_t$	0.093***(0.018)	0.092***(0.018)
$Ln(Capital)_{t-1}$	-0.048***(0.016)	-0.052** (0.016)
$Ln(Employment)_t$	0.040**(0.016)	0.041** (0.016)
$Ln(Employment)_{t-1}$	-0.030*(0.017)	-0.032* (0.018)
$Ln(Energy)_t$	0.179***(0.026)	0.178***(0.026)
$Ln(Energy)_{t-1}$	-0.081***(0.021)	-0.090***(0.022)
$Ln(Materials)_t$	0.614***(0.023)	0.614***(0.023)
$Ln(Materials)_{t-1}$	-0.305***(0.028)	-0.337***(0.037)
$Ln(Highways)_t$	0.029(0.060)	0.033 (0.060)
$Ln(Highways)_{t-1}$	0.151***(0.049)	0.154***(0.047)
$Ln(Employment\ Density)_t$	0.015(0.084)	0.014 (0.083)
$Ln(Employment\ Density)_{t-1}$	-0.069(0.076)	-0.065 (0.074)
Cohort FE	Yes	Yes
Time FE	Yes	Yes
AR1	-	0.000
AR2	-	0.282
Hansen	-	0.000
Number of Instruments		825
Observations	3488	3488
R^2	0.976	-

(where ***, **, * indicate significance at the 1%, 5% and 10% level respectively. Robust standard errors in parenthesis)

Table 5

Robustness Test II – Road Density

Dependent Variable: Ln(Output)$_t$	Road Density	
	Fixed Effects	System GMM
	(1)	(2)
Ln(Output)$_{t-1}$	0.486*** (0.037)	0.538*** (0.052)
Ln(Capital)$_t$	0.093*** (0.018)	0.092*** (0.018)
Ln(Capital)$_{t-1}$	-0.047** (0.016)	-0.051*** (0.016)
Ln(Employment)$_t$	0.040** (0.016)	0.040** (0.016)
Ln(Employment)$_{t-1}$	-0.029* (0.017)	-0.032* (0.018)
Ln(Energy)$_t$	0.180*** (0.026)	0.179*** (0.026)
Ln(Energy)$_{t-1}$	-0.082*** (0.021)	-0.091*** (0.022)
Ln(Materials)$_t$	0.615*** (0.023)	0.615*** (0.023)
Ln(Materials)$_{t-1}$	-0.306*** (0.028)	-0.338*** (0.036)
Ln(Road Density)$_t$	0.021 (0.056)	0.024 (0.052)
Ln(Road Density)$_{t-1}$	0.102** (0.04)	0.104** (0.041)
Cohort FE	Yes	Yes
Time FE	Yes	Yes
AR1	-	0.000
AR2	-	0.332
Hansen	-	0.000
Number of Instruments	-	771
Observations	3490	3490
R^2	0.984	-

(where ***, **, * indicate significance at the 1%, 5% and 10% level respectively. Robust standard errors in parenthesis)

Table 6

Results for Heavy and Light Industries

Dependent Variable: Ln(Output)$_t$	Light Industries		Heavy Industries	
	Fixed Effects	System GMM	Fixed Effects	System GMM
	(1)	(2)	(3)	(4)
Ln(Output)$_{t-1}$	0.442***(0.058)	0.445***(0.071)	0.523***(0.039)	0.552***(0.052)
Ln(Capital)$_t$	0.105***(0.027)	0.105***(0.026)	0.080***(0.018)	0.080***(0.018)
Ln(Capital)$_{t-1}$	-0.060***(0.023)	-0.105***(0.026)	-0.031(0.020)	-0.033*(0.020)
Ln(Employment)$_t$	0.039*(0.020)	0.039*(0.020)	0.035(0.025)	0.036(0.024)
Ln(Employment)$_{t-1}$	-0.039*(0.022)	-0.040*(0.022)	-0.007(0.027)	-0.008(0.027)
Ln(Energy)$_t$	0.198***(0.037)	0.198***(0.037)	0.154***(0.033)	0.154***(0.033)
Ln(Energy)$_{t-1}$	-0.079***(0.029)	-0.079***(0.029)	-0.078***(0.028)	-0.154***(0.033)
Ln(Materials)$_t$	0.617***(0.029)	0.617***(0.029)	0.624***(0.033)	0.625***(0.033)
Ln(Materials)$_{t-1}$	-0.277***(0.042)	-0.279***(0.050)	-0.337***(0.035)	-0.355***(0.043)
Ln(Highways)$_t$	0.007(0.063)	0.007(0.063)	0.043(0.132)	0.041(0.130)
Ln(Highways)$_{t-1}$	0.089**(0.035)	0.086**(0.035)	0.343***(0.106)	0.337***(0.103)
Cohort FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
AR1	-	0.000	-	0.00
AR2	-	0.237	-	0.477
Hansen	-	0.000	-	0.000
Number of Instruments	-	557	-	539
Observations	1853	1853	1652	1652
R^2	0.984	-	0.959	-

(where ***, **, * indicate significance at the 1%, 5% and 10% level respectively. Robust standard errors in parenthesis)

References

ANI, A.N. de I. (2013), 'Autopistasa 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 October 10, 2015].

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

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

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

Baldwin, R.E. & Okubo, T. (2006), 'Heterogeneous firms, agglomeration and economic geography: spatial selection and sorting', Journal of Economic Geography, 6(3), pp.323–346.

Barnes, G. & Langworthy, P. (2003), 'The Per-mile Costs of Operating Automobiles and Trucks', Minnesota Department for Transportation, Minneapolis, M.N..

Blundell, R. & Bond, S. (2000), 'GMM Estimation with persistent panel data: an application to production functions', Econometric Reviews, 19(3), pp. 321–340.

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

Bougheas, P., Demetriades, P. & Mamuneas, T. (2000) 'Infrastructure, specialization and economic growth', Canadian Journal of Economics, 33(2), pp. 506 - 522.

Calderón, C. & Servén, L. (2010), 'Infrastructure in Latin America', Policy Research Working Paper 5317, The World Bank, Washington, D.C.

Calderón, C. & Servén, L. (2004a), 'The Effects of Infrastructure Development on Growth and Income Distribution', Policy Research Working Paper 3400, The World Bank, Washington, D.C.

Calderón, C. & Servén, L. (2004b), 'The Output Cost of Latin America's Infrastructure Gap', in W. Easterly & L. Servén (eds), *The Limits of Stabilization: Infrastructure, Public Deficits, and Growth in Latin America*, The World Bank and Stanford University Press: Stanford, CA.

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

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

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

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

Duranton, G. & Puga, D. (2004), 'Micro-Foundations of Urban Agglomeration Economies', in V. Henderson & J.-F. Thisse (eds), *Handbook of Regional and Urban Economics*, Elsevier B.V. :Amsterdam, Nord-Holland.

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

Faber, B. (2014), 'Trade Integration, Market Size, and Industrialization: Evidence from

China’s National Trunk Highway Syste‘, The Review of Economic Studies, 81(3), pp. 1046-1070.

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

Gimenez-Nadal, J.I. & Molina, A. (2011), ‘Commuting Time and Labour Supply: A Causal Effect?‘, Discussion Paper No. 5529, Institute for the Study of Labor, Bonn.

Graham, D.J. (2007), ‘Agglomeration Economies and Transport Investment, Discussion Paper No . 2007-11, OECD/ITF Joint Transport Research Centre, The OECD.

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

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

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

Infrastructure Journal (2012), IJ Infrastructure Investment Guide: Colombia. London

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

Lahr, M.L., Duran, R. & Varughese, A. (2005), ‘Estimating the Impact of Highways on Average Travel Velocities and Market Size‘, Urban/Regional EconWPA 0403009, Center for Urban Policy Research, Rutgers University.

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

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

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

Nickell, S. (1981), ‘Biases in Dynamic Models with Fixed Effects‘, Econometrica, 49(6), pp. 1417-1426.

Planeacion, D.N. de, 2014. Plan Nacional de 2010 - 2014, available at: https://www.dnp.gov.co/Plan-Nacional-de-Desarrollo/PND-2010-2014/Paginas/Plan-Nacional-De-2010-2014.aspx.

Rodríguez-Pose, A., Tselio, V., Winkler,.D. & T. Farole (2013), 'Geography and the Determinants of Firm Exports in Indonesia', World Development, 44, pp. 225 - 240.

Tripathi, S, & Gautam, V. (2010), ‘Road Transport Infrastructure and Economic Growth in India‘, Journal of Infrastructure Development, 2(2), pp. 135-151.

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

Acknowledgements

We are particularly grateful to the help of Germán Ospina, who is an independent transportation consultant in Colombia, for the data access. We thank the generous help of the World Bank, in particular Antonio Estache, Daniel Alberto Benitez and Felipe Targa for the insight into the Colombian transportation sector. Special thanks to Javier Orlando Aguillon Buitrago at the Colombian Ministry of Transportation and Pablo Roda for the insightful knowledge on transportation in Colombia. We further thank Steve Bond for his insight into dynamic panel models.