Roads and Trade in Colombia

Gilles Duranton[*‡]
University of Pennsylvania

September 2014

ABSTRACT: I estimate the effect of major roads within and between cities on the level and composition of trade for Colombian cities. I confirm that road distance between cities is a major impediment to trade. In addition, major roads within cities have a large effect on a city's exports and imports with an elasticity of approximately 0.20 estimated with OLS and up to 0.50 with IV. If anything, the effects are stronger for the value than for the weight of exports. I interpret these results as city roads shifting economic activity in cities towards the production of tradable and somewhat lighter goods.

Key words: roads, trade, Colombia, specialisation, market access

JEL classification: F14, R41, R49

[*]Thanks to Prottoy Akbar for terrific research assistance and Magda Biesiada for equally wonderful GIS work. I am grateful to Francisco Perdomo, Pablo Roda, and Jose Antonio Pinzon for providing me with the data and for their initial help with them. Thanks also to Peter Morrow, Matt Turner, two anonymous referees, and the editor, André de Palma, for helpful feedback and suggestions. Financial support from the Colombian Ministry of Planning (DNP) and UN-Habitat is gratefully acknowledged. Finally this paper is dedicated to Richard Arnott. No-one has contributed more than Richard to bring serious economic modelling into research on transport. Richard has also broadened the analysis of transport by introducing new topics. Although Richard never worked on the trade-transport nexus that is explored here, I was certainly guided by Richard's example regarding the use of serious economic modelling and expanding the horizon of transport research.

[‡]Wharton School, University of Pennsylvania, 3620 Locust Walk, Philadelphia, PA 19104, USA (e-mail: `duranton@wharton.upenn.edu`; website: `https://real-estate.wharton.upenn.edu/profile/21470/`).

1. Introduction

Road investments are often extremely large. Over 2006-2011, transport represented 16% of World Bank lending, that is about 30 billion dollars. For the US, Allen and Arkolakis (2014) compute the annualised cost of the Interstate Highway System at 100 billion dollars or 0.7% of US GDP. Despite the importance of the amounts at stake, very little is known about the economic effects of road networks.

While roads may affect a broad variety of outcomes, imports and exports to other places are of particular importance. That new or improved roads between cities should foster trade is the main direct consequence that we anticipate from this type of investment. In turn, more trade is often assumed to lead to more employment, higher wages, and other desirable outcomes. To make sure that new roads positively affect economic activity, it is important to verify that they have the intended effect on the main direct outcome, trade, before looking at more indirect outcomes such as wages or employment.

To make progress on this research agenda, this paper investigates the effects of major roads between and within cities on the weight and value of bilateral trade between Colombian cities following the same approach as Duranton, Morrow, and Turner (2014). I first confirm that road distance between cities is a major impediment to trade with an elasticity of trade with respect to distance of around -0.60. Roads within cities also affect trade. A 10% increase in major roads within a city is associated with about a 2% increase in exports to other cities using ordinary least square. Estimates using instrumental variables suggests even stronger effects with a 10% increase in major roads in a city causing up to a 4% increase in exports. There is evidence that the effect of city roads is modestly stronger for the value of trade than for its weight. I interpret these results as roads shifting economic activity in cities at the extensive margin towards goods that are tradable and goods that are lighter (e.g., light manufacturing goods exported to other cities instead of processed food sold locally). These headlines results contrast with those of Duranton *et al.* (2014) where highways in the United States are found to affect trade at the intensive margin and foster the weight of exports but not their value. However, these differences between Colombia and the US are significantly reduced when similar range of distances are considered for trade.

Although there is some emerging evidence about the effects of transport infrastructure on trade, this evidence concentrates either on rich economies with abundant infrastructure such as the early

21st century United States (Duranton *et al.*, 2014) or, at the opposite end, extremely poor economies with no infrastructure such as 19th century India (Donaldson, 2014). Very little is known about middle-income, developing countries such as Colombia where the paucity of infrastructure is often deemed to be a brake on development.[1] More generally, trade is a particularly important outcome since it is often alleged to be a key driver of growth and development (e.g., Redding and Venables, 2004, Alcalá and Ciccone, 2004).

Second, the tendency in the recent literature on the effects of transport infrastructure has been to look at different outcome variables and develop new methodologies.[2] To assess the effects of the US interstate system, Fernald (1999) used aggregate data and essentially performed a difference-in-difference exercise looking at the changes in the productivity of sectors of economic activity depending on their transport intensity. Chandra and Thompson (2000) focus on a variety of outcomes in rural US counties for which the incidental presence of a section of an Interstate Highway may be taken as exogenous. Baum-Snow (2007) and Michaels (2008) pioneered the use of instrumental variables to look at the effects of transport. They focus on suburbanisation and the demand for skills, respectively. This type of approach, which shares some commonalities with the work done here, has been pursued by Duranton and Turner (2012) and Duranton *et al.* (2014) to examine the effects of the US Interstate Highway System on urban growth and trade. More recently, Gibbons, Lyytikäinen, Overman, and Sanchis-Guarner (2012) have developed a spatial differencing approach using geocoded data and fine changes in accessibility to measure the effects of new roads on firm employment.

While it is important to develop new and better approaches and examine different outcomes, there is also a lot of value in taking the same outcome and replicating existing methodologies for different countries. This enables us to gauge the robustness of extant methodologies beyond the specific data for which they were originally developed. This also allows for an assessment of the external validity of the original findings. In this spirit, this paper essentially replicates the US work of Duranton *et al.* (2014) using Colombian data.

Finally, the results of this paper also matter for Colombian public policy. For many years, the Colombian government has somewhat neglected its road infrastructure. This neglect might seem

[1] To the best of my knowledge, the only other piece of research that looks on the effects of roads on trade is Volpe-Martincus, Carballo, and Cusolito (2012) who examine the effects of roads in Peru on firms' exports to other countries.

[2] See Anas, Arnott, and Small (1998) for a broad discussion of the effect of transport on cities. The recent literature is reviewed by Redding and Turner (2014).

surprising in light of a legacy of poor connectivity within the country due to its harsh geography (Bushnell, 1993, Samad, Lozano-Gracia, and Panman, 2012). This relative lack of interest in the transport infrastructure is the consequence of both persistent civil unrest in many parts of the country and other policy priorities such as education and health taking precedence (Barco, 2013). With these policy objectives close to being fulfilled and a major decrease in civil unrest, the country has started to invest in its roads again (Roda, Perdomo, and Sánchez, 2012). Trying to understand whether these on-going investments are appropriate and how they might effect Colombian cities is an important objective for this paper.

Like Duranton *et al.* (2014), my analysis of how roads in Colombia affect trade flows between cities consists of three components: a theoretical framework, which clarifies the effects of roads on trade, motivates the econometric specification, and highlights some identification issues; high quality data describing roads, trade flows and relevant covariates; and a strategy for resolving the possible endogeneity of roads to trade flows. The theoretical model is an extension of Anderson and van Wincoop's (2003) framework to an economy with multiple sectors of production. It follows the model developed by Duranton *et al.* (2014) to which it adds a sector producing a local non-traded good.

This model leads to a gravity equation describing the effects of distance on equilibrium trade flows. It also implies that cities with a relative abundance of roads also export and import more. Finally, this model provides a logically consistent framework to examine the effect of within- and between-city roads on the weight and value of inter-city trade and to assess the effect of roads on the composition of production and trade for a given city.

This model implies a two-step estimation strategy resembling Redding and Venables (2004). The first step estimates a gravity model for the value and weight of trade flows, each as a function of distance and exporter- and importer-specific fixed effects. These fixed effects measure a city's propensity to export (or import) value (or weight) conditional on distance and trading partner characteristics. In the second step, these fixed effects are regressed on within-city roads and other city characteristics.

My first main sources of data is the 2011 Commodity Flow Survey for Colombia which tracks truck shipments on major Colombian roads. These data allow me to calculate bilateral domestic trade flows for a cross-section of Colombian cities. To measure distances between these cities and roads within them, I use a combination of recent maps of Colombian roads and bilateral travel time and travel distance information extracted from Google Maps. I also use a rich set of city level control variables.

It is possible that the desire to trade causes cities to build roads, or that some unobserved city characteristics cause both trade and the availability of roads. Such endogeneity and missing variable problems may confound estimates of the relationship between roads and trade. Resolving these inference problems is the third part of our analysis. Like Duranton *et al.* (2014), I rely on instrumental variables estimation using instruments based on maps of colonial routes ('caminos reales') and the 1938 road network.

The rest of this paper is organised as follow. Section 2 describes the data. Section 3 exposes the model and derives an econometric specification. The estimation issues raised by my empirical strategy are also discussed in this section. Section 4 presents the main results regarding the effects of roads between cities on trade. Section 5 reports OLS and IV results regarding the effects of within-city roads on the propensity of Colombian cities to export and import. This section also reports results for a number of robustness checks. Finally, section 6 concludes.

2. Data

The main source of data is the 2011 Commodity Flow Survey (CFS) for Colombia ('Encuesta Origen - Destino a Vehìculos de Carga') from the Colombian Ministry of Transport. This is a survey of trucks on major Colombian roads. Trucks are pulled aside at weigh stations for inspection and taxation. During the period of the survey, the origin, destination, and the weight and nature of the cargo of trucks passing through a weight station is recorded. Trucks drivers are also asked about the frequency with which they do this particular journey. The objective of the survey is to inform the Colombian Ministry of Transport about freight traffic on major roads in the country.

In the data, an observation is an origin municipality, a destination municipality, a cargo weight, and a product (among 39). The original data contains 159,439 observations. Note that a shipment containing two different products is represented with two different observations. In most of the empirical work below, these observations are aggregated by pairs of municipalities to generate a measure of bilateral trade flows in weight. The Ministry of Transport also proposes average values for these products so that a measure of trade flows in value can be generated. Table 1 offers some descriptive statistics at the municipal level while table 2 reports descriptive statistics for an aggregation into 16 product groups. Note that, although the main regressions below are estimated

Table 1: Descriptive statistics for 902 municipalities

Variable	Mean	Median	Std Dev	Minimum	Maximum
Weight of exported shipments	129.27	5.90	822.91	0	12,592.20
Value of exported shipments	416.08	23.60	2,450.43	0	33,834.30
Weight of imported shipments	124.92	6.90	947.89	0	16,560.60
Value of imported shipments	402.10	19.70	3,022.19	0	48,214.80
Road distance of exported shipments	319.51	292.43	206.83	6.40	1,651.55
Road distance of imported shipments	312.71	270.45	207.87	11.31	1,376.22
Export share of trade, value	0.51	0.51	0.28	0	1.00
Export share of trade, weight	0.47	0.46	0.26	0	0.99
Share of export transactions with empty shipments	0.30	0.29	0.24	0	1
Share of import transactions with empty shipments	0.26	0.24	0.21	0	1
Road frequency, 2012	0.99	1.00	0.85	0	5.00
Road exits, 2012	2.06	2.00	1.91	0	16.00
Length of roads, 2012	16.46	9.81	23.45	0	255.26
Road index, 2012	3.97	3.76	1.53	-3.24	8.13

Notes: Values are in millions of dollars. Weights are in thousands of tons. Lengths and distances are in kilometres. Export share of trade is the export divided by the total (import + export) trade. Road frequency is the number of major roads going through a municipality. Road exits is the number of major road exits out of a municipality. Length of roads is the mileage of major roads going through a municipality. Road index is a function of the logs of road frequency, road exits and road lengths.

for fewer (larger) municipalities, these descriptive statics concern an extensive set of Columbian municipalities.

Although its name is similar, this survey should be distinguished from its US equivalent used, among others, by Duranton *et al.* (2014). The US CFS surveys firms about their shipments whereas the Colombian CFS surveys trucks along the main roads of Colombia. Obviously, the Colombian survey is restricted to road shipments. This is not an issue for Colombia since an overwhelming share of internal trade in Colombia goes by road. Trade along the main river, the Magdalena river, is limited; railroad trade is concentrated on mainly one link for commodities after most railroads were decommissioned years ago; and air shipments are unimportant in most sectors (Roda *et al.*, 2012).

A more significant worry is that short shipments will be under-represented in the data. While long-distance shipments need to use major roads along which weigh stations are located, short distance shipments between neighbouring municipalities are likely to use less important roads and may not be surveyed. This selection bias may lead to biased estimates. I devote considerable attention to this issue below.

Table 2: Descriptive statistics for sectors

Product Group	Value	Weight	$\frac{\text{Weight}}{\text{Value}}$	Routes	Exporters	Importers	Total	Distance
Biofuels and Fertilizers	0.5	1.1	2.14	309	81	129	160	356.5
Cattle, Milk and Animal Oils and Fats	21.3	9.2	0.43	3300	539	536	647	306.6
Cement, Lime, Plaster and Ceramic	24.4	12	0.49	1383	218	422	464	350.6
Chemicals and Non-Metallic Minerals	127	14.8	0.12	1906	370	454	538	255.7
Coffee and Beverages	41.0	14.1	0.34	1585	371	416	498	291.9
Corn, Potato and Rice	18.1	14.5	0 .80	1493	301	348	441	513.6
Food and Animal Feed	33.7	14.4	0.43	2849	458	570	654	325.4
Fruits, Vegetables and Flowers	21.4	5.0	0.23	1779	396	412	526	364.6
Manufactured Goods and Vehicles	17.3	16.5	0.95	2061	360	473	544	415.0
Manufacturing Machinery	39.6	6.6	0.17	1186	263	323	399	446.7
Paper, Cardboard and Textiles	15.8	6.8	0.43	1106	246	285	368	447.2
Petroleum, Coal and Crude Oil	142.9	32.6	0.23	2135	396	497	576	291.6
Rubber, Plastic and Leather	7.3	3.3	0.45	799	205	246	315	446.3
Salt, Soy, Sugar, Wheat and Flours	17.7	12.5	0.71	1114	198	307	367	455.1
Precious woods	71.9	4.5	0.06	1404	339	326	459	341.2
Others	11.3	4.9	0.43	1340	324	340	440	401.6

Notes: 14 groups of originally 39 sectors. Value is in billions of dollars. Weight is in millions of tons. Routes is the number of different municipality trade partners. Total is the total number of trading municipalities. Distance is the average driving distance in kilometres between trading municipalities.

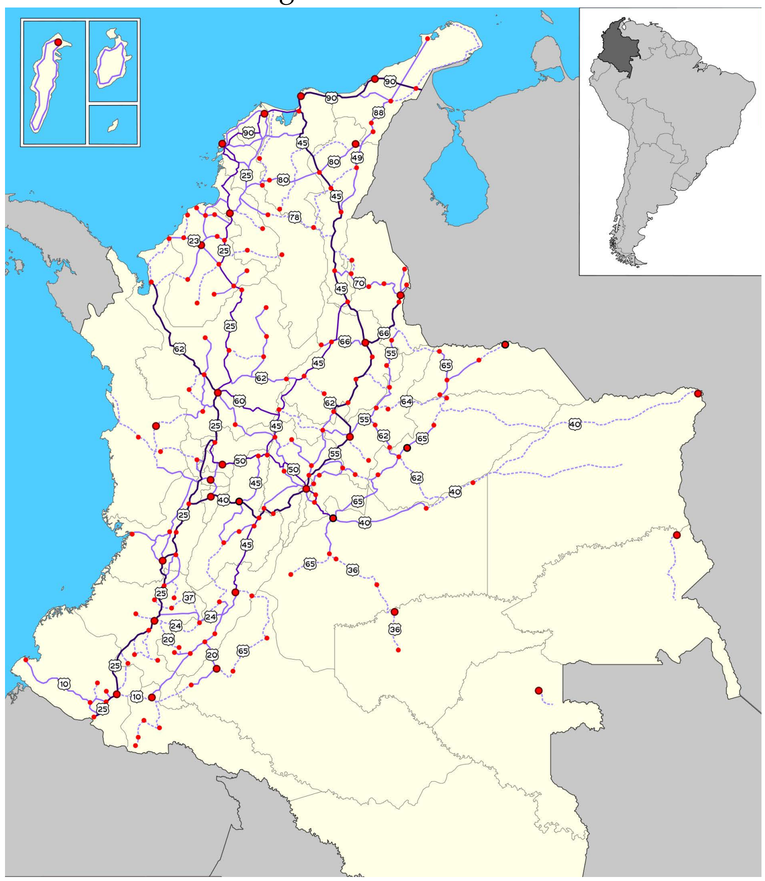
To measure roads, I use several sources. The first source for current roads is a 2010 map from the Colombian Ministry of Transport represented in figure 1. Four types of roads compose the national road network, which I refer to as 'major roads', the Panamerican (the darker thicker plain line on the map), principal roads (dark plain lines), secondary road (light lines), and other roads (dotted lines). There are in total about 16,300 kilometers of major roads in Colombia, which are concentrated along a North-South middle corridor where most of the Colombian population lives.

After converting figure 1 to a digital map with the same format and projection as the map of Colombian municipalities, I computed for each municipality the number of roads, the number of entries, and the number of kilometers of roads. Some municipalities do not have direct access to a major road. For these, I also computed the distance between the municipality centroid and the closest major road.

I supplemented the map in figure 1 with another, coarser map containing information about the number of lanes of roads (one or two except in the central part of the largest cities), existing railroads, and the main river link along the Magdalena river.

While it is possible in theory to use the map of figure 1 to compute distances between municipalities, doing so is would be imprecise and problematic in the case of municipalities with no direct

Figure 1: 2010 roads



Source: Colombian Ministry of Transport.
Notes: Darker thicker plain lines: Panamerican road, dark plain lines: principal roads, light lines: secondary road, dotted lines: other roads.

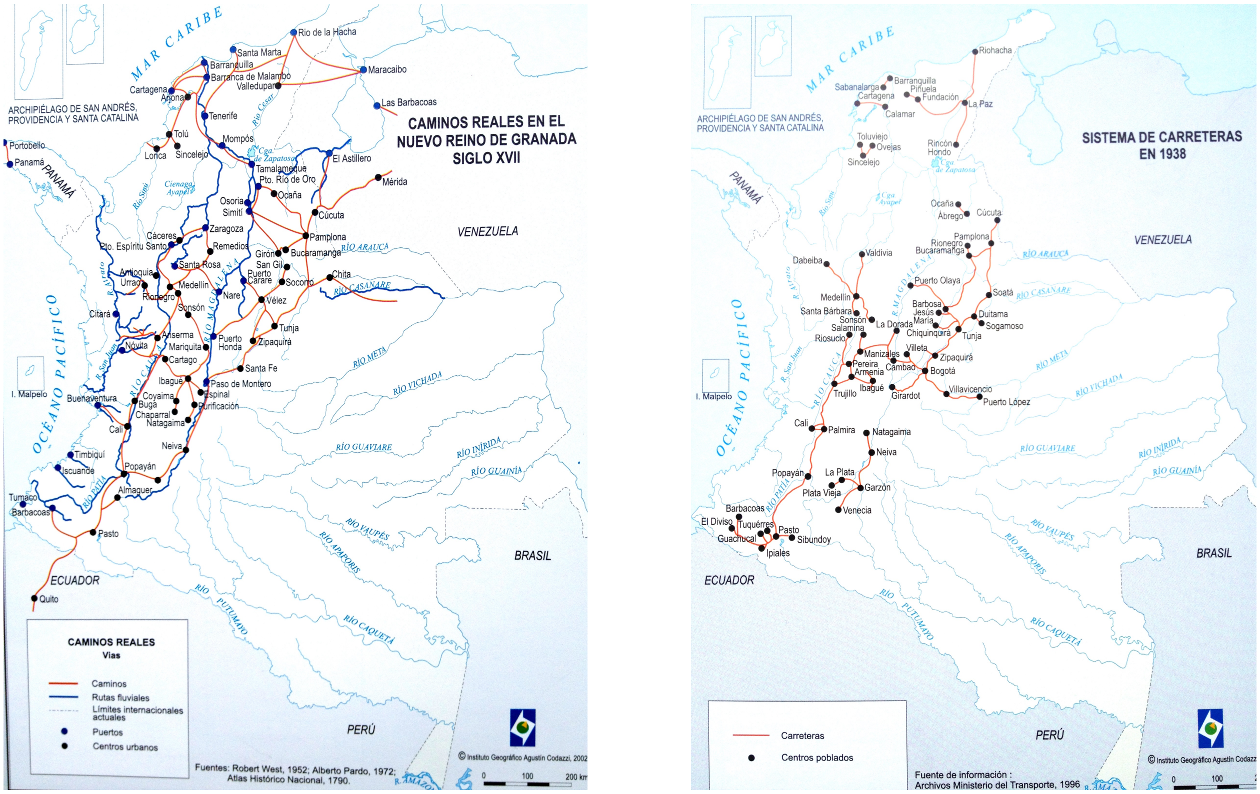
access to a major road. Instead, travel time and road distance data were collected from Google Maps between the centroids of all pairs of trading municipalities. Finally, Euclidian distances between municipalities can be readily computed from their coordinates.

Because contemporaneous roads may be assigned to municipalities on the basis of their propensity to trade (or some correlates of it), the analysis below instruments contemporaneous roads with measures of historical roads, colonial roads from the 17$^{\text{th}}$ century and earlier ('caminos reales' or royal routes) and the 1938 road network. Panel A of figure 2 represents the network of colonial roads while panel B of the same figure represents the 1938 roads network.

Just like with contemporaneous roads, these two maps were converted to a digital map with the same format and projection as the map of Colombian municipalities. Then, the same measures of roads (frequency, length, exits, and distance to the nearest road) were computed.

Figure 2: Historical roads

Panel A: Colonial routes | Panel B: 1938 roads



Source: Colombian Institute of Geography (IGAC — Instituto Geogràfico Augustin Codazzi.)

Finally, I also use a variety of other municipal data. These include population data from the 1951, 1964, 1985, and 1993 censuses as well as recent 2010 population estimates, all from the Colombian Statistical Institute (DANE). As in Duranton (2013), I compute municipal wages net of individual characteristics using the 2011 National Household Survey (2011 GEIH).

3. Theory and estimation

3.1 *Economic model*

Because it is not immediately clear how local roads affect trade, which empirical specification to use, and what are its possible pitfalls, a model is needed. I follow the model proposed by Duranton *et al.* (2014) itself a variant of Anderson and van Wincoop (2003) with Armington preferences tailored to assess the issue at hand. There are two main differences with Duranton *et al.* (2014). First, I add a non-tradable sector that helps the interpretation of the results. Second, to ease the exposition, I consider a continuum of cities rather than a large discrete number.

There is a continuum of cities on the interval [0,1]. City i hosts L_i identical workers, N_i of which are employed in a continuum [0,1] of formal sectors producing tradable goods. The remaining $L_i - N_i$ workers are employed in the informal sector and only consume goods from this sector. We leave the informal sector aside for now but note that this separation between formal and informal sectors is extreme. Its main advantage is to avoid complicated interactions between sectors that cannot be examined empirically with the data used below. In each city i and (formal) sector k, competitive firms produce under constant returns to scale and productivity A_i the same city-specific variety with N_i^k workers:

$$Q_i^k = A_i \, N_i^k \,. \tag{1}$$

A consumer in city j that works in a formal sector maximizes utility,

$$U_j = \left[\int_0^1 \int_0^1 (q_{ij}^k)^{\frac{\sigma-1}{\sigma}} \, \mathrm{d}i \mathrm{d}k \right]^{\frac{\sigma}{\sigma-1}} \qquad (\sigma > 1) \,, \tag{2}$$

with respect to consumption q_{ij}^k of the varieties from all cities and sectors subject to the consumer's budget constraint,

$$W_j \geq \int_0^1 \int_0^1 P_{ij}^k \, q_{ij}^k \, \mathrm{d}i \, \mathrm{d}k \,, \tag{3}$$

where W_j is the wage in j and P_{ij}^k is the price in city j of the variety from city i in sector k.[3] Solving for the consumer programme and aggregating across consumers yields the value of shipments from i to j in sector k:

$$X_{ij}^k \equiv P_{ij}^k \, Q_{ij}^k = \left(\frac{\mathbb{P}_j}{P_{ij}^k} \right)^{\sigma-1} W_j \, N_j \qquad \text{with } \mathbb{P}_j \equiv \left[\int_0^1 \int_0^1 (P_{i'j}^{k'})^{1-\sigma} \, \mathrm{d}i' \mathrm{d}k' \right]^{\frac{1}{1-\sigma}} \tag{4}$$

where Q_{ij}^k is the exported quantity from i to j in sector k and $\mathbb{P}_j$ is the price index for city j.

Trade between cities is costly. For one unit of output from city i's sector k to reach importer city j, the exporter must ship $\tau_{ij}^k \geq 1$ units. This is the standard 'iceberg' specification that is widely used in international trade. With marginal cost pricing by firms, prices are equal to $P_{ij}^k = \tau_{ij}^k \frac{W_i}{A_i}$. Using this expression into equation (4) and aggregating across sectors yields the value of aggregate shipments from i to j:

$$X_{ij} = \left(\frac{A_i}{W_i} \right)^{\sigma-1} \left[\int_0^1 \left(\tau_{ij}^k \right)^{1-\sigma} \mathrm{d}k \right] \mathbb{P}_j^{\sigma-1} \, W_j \, N_j \,. \tag{5}$$

[3] The right-hand sides of equations (2) and (3) integrate over the continuum of sectors and variety. Having sectors noted as superindices and locations as subindices is an abuse of notation that allows these expressions to remain of manageable length and readability.

To derive some predictions and be able to put this expression to the data, three further steps are needed. First, the model needs to be solved further for the labour market equilibrium among formal sectors in each city. The demand for labour in city i and sector k is readily obtained from the production function. Then, using the price expression above into equation (4) and recalling that τ_{ij}^k units must be shipped from city i for unit to arrive in city j and for Q_{ij}^k units to be consumed in city j yields an expression for equilibrium employment in city i and sector k. Aggregating across sectors yields, after simplification, the aggregate demand for labour in formal sectors,

$$N_i = \frac{A_i^{\sigma-1}}{W_i^{\sigma}} \int_0^1 \int_0^1 \frac{\mathbb{P}_j^{\sigma-1}}{(\tau_{ij}^k)^{\sigma-1}} W_j N_j \, \mathrm{d}j \mathrm{d}k \, . \tag{6}$$

Second, we need to specify labour supply in formal sectors. While there would be many possible microfoundations to justify duality in the labour market, I assume that the supply of labour in formal sectors is given by

$$N_i = \left(\frac{W_i}{\overline{W}} \right)^{\gamma} L_i \tag{7}$$

if the wage in the formal sectors is below a reservation threshold $\overline{W}$ and $N_i = L_i$ otherwise. The parameter γ is the elasticity of labour supply in the formal sectors with respect to the wage in those sectors.

Third, more structure is needed for transport costs. Assume: $\tau_{ij}^k \equiv \tau_x^k(R_i) \times \tau_{xm}(R_{ij}) \times \tau_m(R_j)$. That is, the transport cost in sector k between city i and j has three parts: the cost of leaving i, $\tau_i^k \equiv \tau_x^k(R_i)$, which depends on roads in this city R_i; the cost of going from i to j, $\tau_{ij} \equiv \tau_{xm}(R_{ij})$, which depends on distance between these two cities R_{ij}; and the cost of entering j, $\tau_j \equiv \tau_m(R_j)$, which depends on roads in this city R_j. Although they are noted similarly, roads variables have slightly different meanings: R_{ij} refers to a distance between two cities whereas R_i and R_j refer to a measure of roadway within a city. While roads are the main determinants of transport costs in this model, disentangling the effects of roads and other city characteristics is the main empirical challenge in the empirical analysis below.

This proposed specification for transport costs implies that the cost of leaving city i depends on the roads within this city: $\partial \tau_x^k(R_i)/\partial \log R_i < 0$ and differs across sectors. After denoting V^k, the weight per unit in sector k, sectors can be ranked from the heaviest to the lightest (i.e., from 0 to 1). It is possible that sectors producing heavier goods are more sensitive to roads. Formally, this would be equivalent to assuming that the elasticity of the transport cost with respect to city roads, i.e., $\rho_R^{\tau,k} \equiv \partial \log \tau_x^k(R_i)/\partial \log R_i$, is larger in absolute value for sectors producing heavier goods.

This assumption may be justified when heavier goods make more bulky shipments which may benefit more from having more and better roads. This assumption need not always be warranted. More roads allowing to exit a city surrounded by mountains (not an unusual case in Colombia) may favour lighter shipments more than heavier shipments for which the steepness of the road may remain a major issue. The purpose of the empirical exercise below is to assess if imports and exports increases with a city's road and which goods, heavy or light, benefit the most.

The above specification for transport also assumes that the cost of transport between cities decreases with roads between cities, i.e., $\partial \tau_{xm}(R_{ij})/\partial \log R_{ij} < 0$. Unlike the cost of exiting cities, there is no variation across sectors in the cost of travelling between cities. Some extensions of the empirical analysis below allow for some heterogeneity across sectors. This is ignored in this model to keep it transparent and easily tractable. Finally, the cost of entering city j decreases with its roads, i.e., $\partial \tau_m(R_j)/\partial \log R_j < 0$. Again, we assume no heterogeneity across sectors in entry costs to avoid the complications associated with patterns of city specialisation being determined by roads in a city's neighbours.

There are four main predictions from this model:

1. A reduction in road distance between two cities increases the value of trade between these two cities but not its composition.

2. An increase in roads within a city increases the wage and employment in formal sectors.

3. An increase in roads within a city increases the value and weight of its exports and imports.

4. An increase in roads within a city increases its relative employment in sectors producing heavy goods and reduces its relative employment in sectors producing light goods when lower exit costs benefit heavy goods more, and conversely.

While the proofs for these results can be found in Duranton *et al.* (2014), their intuitions are fairly straightforward.[4] Lower transport costs caused by a reduction in road distance between cities make all goods produced in these cities less expensive and, through a demand response, this leads to more trade in equilibrium. Because goods have the same elasticity of substitution, a proportional decline in transport costs will affect all goods in the same way and thus will not affect the composition of trade. Turning to roads within cities, an increase in roads within

[4] Duranton *et al.* (2014) do not consider an informal sector but the result about an increase in formal-sector employment follows directly from equation (7).

a city also makes it cheaper for producers in this city to export. Again, through the demand response from other cities, the value of exported goods increases with more roads within a city. In turn, this causes labour demand to increase and wages to go up which leads to an expansion of employment in the formal sectors. Since trade is balanced, more exports must also lead to an increase in imported values. Then, because within-city roads affect sectors differently, the increase in exports is stronger in sectors that benefit the most from this decline in the cost of exiting a city. If sectors producing heavy goods benefit the most, sectors producing lighter goods end up contracting in relative terms.[5] The opposite would hold if sector producing lighter goods see the greatest reduction in trade costs with more roads. Put differently, an increase in city roads can alter the patterns of comparative advantage and cause greater specialisation in the production and export of some types of goods.

Beyond these qualitative predictions, the model can also be used to derive a specification for the regressions estimated below. Using the expression above for trade costs into expression (5), making use of equation (6), taking logs, and reorganising the resulting expression implies

$$\ln X_{ij} = \delta_i^X + (1-\sigma)\ln \tau_{ij} + \delta_j^M\,, \tag{8}$$

where

$$\delta_i^X = S(R_i) + \frac{\sigma-1}{\sigma}\log A_i + \frac{\sigma-1}{\sigma}\log N_i - \frac{\sigma-1}{\sigma} MA_i^X \tag{9}$$

and

$$\delta_j^M = S(R_j) + \frac{\sigma-1}{\sigma}\log A_j + \frac{\sigma-1}{\sigma}\log N_j + \frac{\sigma}{(\sigma-1)^2} MA_j^X - MA_j^M\,. \tag{10}$$

In turn, the terms $S(R_i)$, MA_j^X, and MA_j^M are given by:

$$S(R_i) \equiv \frac{1}{\sigma}\ln\int_0^1 \left(\tau_i^k\right)^{1-\sigma}\,\mathrm{d}k\,, \tag{11}$$

$$MA_i^X \equiv \ln\int_0^1 \frac{\mathbb{P}_j^{\sigma-1}}{(\tau_{ij}\,\tau_j)^{\sigma-1}} W_j\,N_j\,\mathrm{d}j = \log\int_0^1 e^{(1-\sigma)\,\ln\tau_{ij}+\delta_j^M}\,\mathrm{d}j\,, \tag{12}$$

$$MA_j^M \equiv (1-\sigma)(\log\mathbb{P}_j - \log\tau_j) = \log\int_0^1 e^{\delta_i^X+(1-\sigma)\log\tau_{ij}}\,\mathrm{d}i\,. \tag{13}$$

Since N_i, employment in formal sectors, is hard to observe in the data, it is possible to rewrite equations (9) and (10) using equations (5) and (7) as a function of city population L_i,

$$\delta_i^X = \frac{\sigma(1+\gamma)}{\sigma+\gamma}S(R_i) + \frac{(\sigma-1)(1+\gamma)}{\sigma+\gamma}\log A_i + \frac{\sigma-1}{\sigma+\gamma}\log L_i - \frac{\sigma-1}{\sigma+\gamma}MA_i^X - \gamma\frac{\sigma-1}{\sigma+\gamma}\overline{W} \tag{14}$$

[5] If labour supply in formal sectors is constant as in Duranton *et al.* (2014) or if the elasticity of employment in formal sectors with respect to wages, γ is low enough, employment in light sectors will decrease in absolute terms. If γ is large enough employment in light sectors will increase but less than employment in heavy sectors.

and

$$\delta_j^M = \frac{\sigma(1+\gamma)}{\sigma+\gamma} S(R_j) + \frac{(\sigma-1)(1+\gamma)}{\sigma+\gamma} \log A_j + \frac{\sigma-1}{\sigma+\gamma} \log L_j + \left(\frac{\sigma}{(\sigma-1)^2} + \frac{\gamma(\sigma-1)}{\sigma(\sigma+\gamma)} \right) MA_j^X - MA_j^M + \gamma \frac{\sigma-1}{\sigma+\gamma} \overline{W} \,. \tag{15}$$

In expression (8), the value of trade from i to j is written as a function of propensity to export (value) of city i, δ_i^X, the cost of distance between city i and city j, $(1-\sigma)\ln \tau_{ij}$, and the propensity to import of city j, δ_j^M. In turn the propensities to export and import of cities depend on the fundamentals of the model. The propensity to export is determined by the five terms in expression (14). The first is a function of within city roads $S(R_i)$. The second and third are productivity and population in the city. The fourth term, MA_i^X is an export market access term. This term depends on the price index in all cities, the cost of transport between cities, income, and the cost of entry into these cities.[6] Empirically, not all the components of export market access are observable. Fortunately, export market access can also be rewritten as a function of the distance term and the propensities to import as shown by equation (12). Finally, the last term depends on the reservation threshold $\overline{W}$ in the formal sectors.

The propensity to import described by equation (15) depends on the same five terms as the propensity to export plus an import market access term. This import market access term is also a function of a price index and a cost of entry into cities which are not observable. Like with export market access, equation (13) shows that it is possible to express import market access as a function of terms that are estimated in equation (8), the distance term and the propensity to export.

A similar approach can be used for the weight of trade. Recall that for a unit of the variety from sector k to arrive in j, τ_{ij}^k units need to be shipped from city i. The weight of those goods is $\tau_{ij}^k Q_{ij}^k V^k$. Some simple derivations show that the weight of the trade from city i to city j, T_{ij}, is

$$\log T_{ij} = \delta_i^T + (1-\sigma)\log \tau_{ij} + \delta_j^M \,. \tag{16}$$

In turn, city i's propensity to export weight is given by

$$\delta_i^T = S^T(R_i) + \log A_i + \log N_i - MA_i^X \,, \tag{17}$$

[6] That the export market access of a city affects its propensity to export negatively may seem counter-intuitive. To understand why this is nonetheless the case, note that equation (8), which describes δ_i^X, already accounts for distance to importers and their propensity to import. Equation (14) then only captures a negative indirect wage effect. Equation (5) shows that exports from city i depend negatively on the wage of this city whereas wages and market access are positively related in equation (6). Hence, there must be a negative effect of market access on the propensity to export when substituting market access for wages as in expression (14).

or, after replacing N_i by L_i using expression (7),

$$\delta_i^T = \left(1 + \frac{\gamma\sigma}{\sigma+\gamma}\right) S^T(R_i) + \frac{\sigma(1+\gamma)}{\sigma+\gamma}\log A_i + \frac{\sigma}{\sigma+\gamma}\log L_i - \frac{\sigma}{\sigma+\gamma} MA_i^X + \gamma\frac{\sigma}{\sigma+\gamma}\log \overline{W}\,, \quad (18)$$

where

$$S^T(R_i) \equiv \log \int_0^1 \left(\tau_i^k\right)^{1-\sigma} V^k \, \mathrm{d}k - \log \int_0^1 \left(\tau_i^k\right)^{1-\sigma} \mathrm{d}k\,. \quad (19)$$

An analogous expression can be derived to obtain a corresponding expression for the propensity to import weight.

These derivations take city population as given. Duranton *et al.* (2014) consider a simple extension with labour mobility. The corresponding expressions for the propensity to export or import weight or value are very similar to those written above.[7]

3.2 *Econometric model*

Equations (8) and (14)-(15) for trade in value and the corresponding equations for trade in weight suggest a two-step estimation approach. The first step regresses bilateral trade flows on an exporter fixed effect, a measure of the cost of trade between the exporter and the importer, and an importer fixed effect. These two fixed effects estimate the propensity of cities to export and import. Note that theory provides no guidance about the exact functional form to use for the distance term. In the regression below, I approximate the road distance term of the model R_{ij} with the time cost of travel and estimate the following standard "gravity" regression for trade in value:

$$\log X_{ij} = \delta_i^X + \alpha \log R_{ij} + \delta_j^M + \epsilon_{ij}\,. \quad (20)$$

I also estimate a similar regression for trade in weight corresponding to equation (16) in the model. Since the log of the time cost of travel may not provide the best approximation for the transport cost between i and j, I experiment with a variety of alternative measures of distances and functional forms. Because the time cost of travel and trade flows may be simultaneously determined, I also implement a simple instrumental variable strategy.

Then, at the second step, the propensity to export (or import) is regressed on city roads, city population, and market access. Note that by equations (12) and (13), the two market access

[7]The main difference is that under labour mobility the exporter effect no longer contains productivity but also contain importer market access. Since productivity is not well observed, the first difference is that a simpler error term is expected under labour mobility. It is wise to retain the possibility of a more complicated error term in the empirical analysis. The second difference is empirically unimportant since exporter and importer market access are highly correlated and thus hard to separate empirically.

variables can be computed directly from the output of the first step.[8] Hence, equation (14) maps directly into the following regression:

$$\hat{\delta}_i^X = \beta_0 + \rho_R^X \log R_i + \beta' C_i + \mu_i \,, \tag{21}$$

where the dependent variable is estimated in equation (20), C_i is a vector of city i characteristics and the coefficient on within city roads, ρ_R^X, can be interpreted as an elasticity of exports with respect to city roads.

A similar regression can be estimated for trade in weight using the fixed effects estimated in the analog to regression (20) for trade in weight. This regression is the empirical counterpart to equation (14) in the model. Note finally that the importer fixed effect estimated in equation (20) can also be used as dependent variable and regressed on city roads, population, and import and export market access as suggested by equation (15) in the model.[9] Another, similar regression can also be estimated using the importer effect from the analog of equation (20) for weight.

As highlighted by Duranton *et al.* (2014), the error term in regression (21) will contain both unmeasured (or poorly measured) city productivity and some measurement error. This measurement error is caused by the approximation of $S(R_i)$ in the theoretical expression (14) by $\rho_R^X \log R_i$ in regression (21) and by the use of a variable estimated in the first step as dependent variable.[10] In turn, the composition of the error term suggests a number of threats to identification. First, more roads may be built in cities that have a higher (or perhaps lower) propensity to trade. Second, there may be some missing variables in regression (21) that determine both trade and roads. Third, the measurement error in this regression may be correlated with the propensity to export.

3.3 *Instrument validity*

While measurement problems can be dealt with using alternative measures for city roads, problems of missing variables and reverse causation can be addressed by an appropriate instrumental

[8]As made clear by equations (12) and (13), the two market access variables are complex non-linear aggregates of variables that are not observed (transport costs and price indices). However, they can be recovered from the distance coefficient and the fixed effects estimated in the first step. This elegant strategy was originally developed by Redding and Venables (2004).

[9]Since import and export market access are highly correlated, attempting to separate their effects in the empirical counterpart to equation (10) is unlikely to give reliable results. I only use export market access in the regressions below when using the propensity to import as dependent variable.

[10]Using an estimated coefficient as dependent variable may create some heteroskedasticity. This normally calls for a correction of standard errors. On their smaller sample of 66 US regions, Duranton *et al.* (2014) show that their FGLS correction makes minimal difference.

variable strategy. This strategy mirrors that of Duranton *et al.* (2014) and relies on historical road networks.

The first instrument is based on roads from the era of colonisation of the country ('caminos reales'). These routes were taken by the first colonisers trying to penetrate the interior of the country in search of the gold of the El Dorado. Even before that, many of these roads were used as paths by the indigenous population. Until the late 19th century internal trade was minimal.[11] The main Colombian rivers are not easily navigable and colonial roads were in general no more than trails. Only mules could be used. This was extremely slow and costly. As a result, most Colombian regions were living in quasi autarchy. Put differently, colonial roads were developed for reasons unrelated to modern trade, under a very different set of circumstances. For instance, long after caminos reales where established at the time of independence in 1819, the Colombian population was still only between 1 and 2 million, whereas today it is about 47 million.

In 1938, the road network of Colombia represented in panel B of figure 2 was still rudimentary. An important feature of this road network is that it was mostly regional in scope. There was for instance no link between Bogotá and Medellín and none of these two major cities was connected to either coast. The 1938 road network was mostly connecting neighbouring cities located less than 100 kilometres apart. Today, Colombian municipalities trade with other municipalities that are on average located more than 300 kilometres away. In 1938, the Colombian population was still less than 9 million which corresponds to the population of the greater Bogotá today. Until the 1950s, a large majority of the Colombian workforce was employed in agriculture. Its share today is less than 20%. Hence, like caminos reales, 1938 roads have little to nothing to do with contemporaneous trade patterns and were developed under extremely different circumstances.

The bad news here is that these two historical road networks are fairly different from the contemporary road network and this may limit the ability of historical roads to predict contemporary roads in Colombian cities. I pay attention to the issue of weak instruments below.

To be valid, the instruments must also be orthogonal to the error term in the structural regression (21) and its analogs for imports and trade in weight. As already argued, a direct link between early roads and contemporaneous trade is highly unlikely. This said, the Colombian road network may have been determined by the same missing variables that drive today's propensity to trade of Colombian municipalities. To avoid potential indirect correlations between early road networks

[11]See Bushnell (1993) for more on these issues.

and contemporaneous trade going through other city characteristics we use regional indicators and controls for current population as well as past population for 1951, 1964, 1985, and 1993. By controlling for the past populations in logs we control for the long term growth trajectory of these municipalities and thus hopefully for factors that affect their propensity to trade and their road endowments. We also use a measure of current wages in some regressions.

This strategy, like all instrumental-variable strategies relies on the absence of a correlation between the instruments and the error term in the structural regression. Since negative results can never be decisively proven short of a full randomised controlled experiment, it is always open to doubt. Reassuringly, the same strategy in Duranton *et al.* (2014) for US cities is scrutinised thoroughly and the answers provided by this IV strategy turn out to be remarkably robust. Even if this strategy is imperfect, replicating it closely for another country with different data makes the results easily comparable, which is important in its own right.

4. The effects of road on trade: first-step gravity results

Table 3 reports estimates of the elasticity of trade with respect to distance. The estimated specifications correspond to the first-step 'gravity' equation (20) for the value of trade or its corresponding equation for weight. In panel A, the dependent variable is the weight of bilateral trade. Panel B reports corresponding results for the value of bilateral trade. There are marginal differences in the sample of observations between both outcomes caused by a few goods for which shipment values are unknown.

Column 1 reports the results of an OLS regression of trade flows on exporter and importer fixed effects and regular travel time between origin and destination. This last variable is used as measure of distance. Columns 2 and 3 report results for a similar regression using, as measure of distance, road distance and Euclidian (or straight-line) distance, respectively. Column 4 returns to distance measured by travel time and adds a quadratic term to the specification of column 1. Column 5 further introduces a cubic and a quartic term for travel time. Column 6 considers all three measures of distance (travel time, road distance, and Euclidian distance) in the same specification. Column 7 is the two-stage least-square (TSLS) analog of column 1 where travel time is instrumented by Euclidian distance. Finally, column 8 instruments the four powers of travel time in column 5 by the corresponding powers of Euclidian distance. Although not reported in table 3, note that all

Table 3: First-step results

	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) TSLS	(8) TSLS
Panel A. Dependent variable: (log) Weight of bilateral trade flows.								
log(driving time)	-0.62[a]			-1.10[a]	-1.02[a]	-0.080	-0.63[a]	-1.23[a]
	(0.035)			(0.087)	(0.11)	(0.18)	(0.033)	(0.13)
log(road distance)		-0.55[a]				-0.65[a]		
		(0.031)				(0.23)		
log(Euclid. dist.)			-0.52[a]			0.17		
			(0.030)			(0.14)		
log(driving time)^2				0.18[a]	-0.14			-0.22
				(0.029)	(0.10)			(0.20)
log(driving time)^3					0.17[b]			0.28[b]
					(0.073)			(0.14)
log(driving time)^4					-0.025			-0.048[c]
					(0.016)			(0.027)
R^2	0.44	0.44	0.44	0.44	0.44	0.44		
First stage stat.							36,065	9.26
Panel B. Dependent variable: (log) Value of bilateral trade flows.								
log(driving time)	-0.69[a]			-1.17[a]	-1.17[a]	-0.055	-0.71[a]	-1.47[a]
	(0.039)			(0.094)	(0.12)	(0.21)	(0.037)	(0.14)
log(road distance)		-0.61[a]				-0.78[a]		
		(0.034)				(0.26)		
log(Euclid. dist.)			-0.58[a]			0.22		
			(0.034)			(0.16)		
log(driving time)^2				0.17[a]	-0.087			-0.29
				(0.031)	(0.11)			(0.21)
log(driving time)^3					0.18[b]			0.44[a]
					(0.082)			(0.15)
log(driving time)^4					-0.033[c]			-0.087[a]
					(0.018)			(0.030)
R^2	0.40	0.40	0.40	0.41	0.41	0.41		
First stage stat.							35,502	9.17

Notes: All regressions include importer and exporter fixed effects for all cities. Both panels report the same regressions but use different dependent variables: weight of trade flows for panel A and value of trade flows for panel B. Regressions are based on 7,746 observations in panel A and 7,598 observations in panel B. In columns 7 and 8, travel time variables are instrumented by their corresponding Euclidian distance terms. Robust standard errors in parentheses. ^{a}, ^{b}, ^{c}: significant at 1%, 5%, 10%.

columns also estimate two full sets of fixed effects for both exporters and importers. I first discuss the results reported in table 3 before turning to the fixed effects.

In column 1, the elasticity of trade with respect to distance is -0.62 for weight and -0.69 for value.

Column 2 estimates close, but slightly lower, distance coefficients using road distance instead of travel time as explanatory variable. The coefficients on distance are even marginally lower in column 3 where Euclidian distance is used. The results from columns 4 indicate some concavity when a quadratic term is included. This mild concavity is confirmed by the quartic specification of column 5. Column 6 shows that when all three measures of distance are used together, only road distance is significant. Instrumenting travel time in columns 7 and 8 leaves results mostly unchanged. This reflects the high correlation between travel time and Euclidian distance and is suggestive of an absence of a large endogeneity bias regarding travel time.[12] Although unsurprising, these results vindicate the first prediction derived above: trade in weight and in value decreases strongly with distance. Although the distance coefficients for trade in weight and for trade in value often differ in a statistical sense, this difference is economically small, about 10%.

Before commenting these results further it is worth asserting their robustness. Because the data is collected at weigh stations along major roads, there is a possibility that short distance shipments are under-represented in the data. To verify the findings of table 3, table 16 in appendix reproduces the same regression but censors all shipments of less than 100 kilometres. Should smaller distances be under-represented because of selection caused by the data collection process, the coefficient on distance in say panel A and column 1 should be biased downwards. That is, it should be smaller in absolute value in table 3 relative to table 16. If anything the opposite occurs. In results not reported here, I also performed the same exercise again with a higher censoring threshold of 200 kilometres and confirmed these conclusions.

These smaller distance effects when short shipments are not considered appear to be driven mostly by the curvature of the distance effect. Several elements are indicative of this. First, using the coefficients of column 4 in panel A of table 3, the mean distance effect is equal to -0.56 for all shipments and -0.43 for shipments above 100 kilometers. This difference is about the same as the differences in coefficients between tables 3 and 16 for columns 2 or 3. Second, given the curvature in the distance effect, the truncation of shorter shipments should affect the linear estimations in columns 1-3 but make little difference when non-linear distance effects are estimated (provided

[12]Governments may build more direct roads between places that trade more with each other. The high correlation between road distance and Euclidian distance does not support this assertion. On the other hand, Colombian transport policy aims to build better roads between places that trade more (see Barco, 2013, and references therein). But with more traffic along these roads, there appear to be no obvious gains in travel time as evidenced by the high correlation between travel time and Euclidian distance and the fact that instrumenting travel time by Euclidian distance makes no change to the distance coefficient in table 3.

the approximation is good). This is the case, in particular, in column 4 of tables 3 and 16 where the linear and quadratic coefficients are fairly precisely estimated.

Hence the distance effects reported in table 3 are not an artefact of the data collection process.[13] This is important since the distance coefficients in table 3 around -0.50 or -0.60 are at the low end of those collected by Disdier and Head (2008) in their meta analysis of gravity regressions. These estimates are also considerably lower than those of Duranton *et al.* (2014) who perform the same type of analysis using internal trade data for the US. Their estimates for the mean elasticity of trade in weight with respect to distance are between -1.6 and -1.9.

This, perhaps, surprising finding of a lower sensitivity of trade with respect to distance might reflect an attenuation bias associated with measurement problems in Colombia. For instance, trucks might face different traffic conditions depending on the season or might choose routes different from the one assigned by Google Maps. While these are caveats to be acknowledged, they are perhaps less important than in comparable studies for other countries. By virtue of Colombia's location close to the equator, temperatures barely fluctuate. The scarcity of roads in Colombia usually leaves truck drivers with a single option to reach their destination unlike US commuters (Arnott, de Palma, and Lindsey, 1990). Besides, the Google Maps measures of distance used here are better than most and, in any case, are highly correlated with Euclidian distances. Hence, the much lower sensibility of trade to distance in Colombia reflects something real and is extremely unlikely to be spurious.

It is also hard to imagine that the results are driven by a sad Colombian particularity, the drug trade. This trade is illegal and seeks to avoid controls at weigh stations. Hence the results presented here are mostly exclusive of drug (to the extent that other crops and products are not used to hide cocaine). This trade is also quantitatively tiny in weight and small in value, less than 1% of GDP. Historically, drugs nonetheless have had a large effect on internal trade in Colombia through the associated civil unrest (Dube and Vargas, 2013). Although much of the violence that affected trade had disappeared in 2011 when the trade data used here were collected, this is a limitation to the external validity of our results. More violence along major roads in Colombia

[13]One may also be preoccupied by the large number of zeroes in the data. We only observe shipments for around 7,700 bilateral pairs when Colombia is composed of just above 1,100 municipalities, that is potentially more than a million pairs. First, we only observe a small subset of all shipments and many of the zeroes are perhaps shipments that are not observed and there is no formal censoring by the data provider unlike in the US. Second, there is no obvious way to account for selection into trade. It could only be identified from functional form here. Finally, Duranton *et al.* (2014) show that accounting for the zeroes in trade makes little difference on US data.

probably leads to an underestimate of the effects of roads on trade relative to a situation without civil unrest.

It may seem paradoxical that trade within Colombia where geography is difficult and the road network limited should be much less sensitive to distance than trade within the US. A possible key to understand this apparent puzzle is that in equation (8), the coefficient on distance depends on the elasticity of substitution across goods, σ. Hence, a more integrated economy like the US may be trading more substitutable goods over short distances relative to a less integrated economy like Colombia. Comparing across the two panels of table 3 (as well as between the two panels of table 16), it may also seem surprising that trade measured in weight is less sensitive to distance than trade measured in value. While the difference is modest (around 0.06-0.08) it is often significant and extremely persistent across all specifications. This weight-value difference could be explained if less substitutable goods are also more bulky.

To shed some empirical light on this conjecture, table 17 in the appendix re-estimates the specifications of panel B of table 3 for trade in value adding terms for product groups interacted with distance: agricultural goods, processed food, and minerals, with manufacturing being the reference. Although in table 17, the coefficients are only generally significant for agriculture and food, the point estimates for the distance effects for all non-manufacturing product groups are positive.

While I know of no directly comparable estimates of the elasticity of substitution across broad product groups, Broda and Weinstein (2006) report estimates for the elasticity of substitution between 3-digit product groups. Their results point at many low elasticities of substitution for minerals and agricultural goods. That, within a country, foodstuff should be shipped over fairly long distances from regions of production to the cities of consumption may not be very surprising in the end. The same argument applies probably with even greater strength to many minerals. For instance, the distance between Cerrejón, the main coal mine in Colombia, and Barranquilla, the main port city on the Atlantic, is about 350 kilometers. This is more than the median distance of shipments of 226 kilometers.[14] Other major cities are even further away.

Then, the difference in the distance coefficients between weights traded and values traded becomes easy to explain knowing that, on average, the value per ton for manufacturing products is more than 50% larger than the value per ton for other products. It follows that the 'average'

[14] This figure differs from that reported in table 1 which is computed at the municipal level.

distance sensitivity to trade in value is higher than for trade in weight since manufacturing goods, which are more sensitive to distance, get weighted more when trade is measured in value.

As already noted, when all three measures of distance are used together in column 6 of table 3, road distance is the only one that is significant and of the same magnitude as when introduced alone. While it could be tempting to conclude that road distance is indeed the best measure of the cost of distance, it is probably too early to provide a definitive conclusion. The pairwise correlation is 0.965 for road distance and time distance, 0.962 for road distance and Euclidian distance, and 0.905 for time distance and Euclidian distance. As a result small differences in the quality of how road distance vs. travel time are measured could explain this outcome. In addition, there is slightly more variation in road distance than in time distance which could explain the marginally smaller coefficient in column 2 of table 3 relative to column 1.

Turning to the fixed effects estimated in table 3, the main result to keep in mind is that the estimates of the exporter and importer fixed effects are highly stable across specifications. For cities with population above 40,000, the pairwise correlations between the sets of exporter fixed effects estimated in the eight regressions of panel A are all 0.98 or above, including 0.999 between the estimates of exporter fixed effects of columns 1 and 2. The same holds for importer fixed effects. The corresponding correlations for the fixed effects estimated in panel B are equally strong. On the other hand, the correlation within a given specification between exporter and importer fixed effects are much lower, of the order of 0.5 for the regressions of panel A and 0.6 for the regressions of panel B.

The high correlations between the fixed effects estimated in the different specifications of table 3 are reassuring regarding the expected robustness of the findings of the second step. The second-step regressions that follow generally rely on importer and exporter fixed effects estimated in column 1. Robustness checks presented below experiment with importer and exporter fixed effects based on other first step specifications.

Finally, note that the R^2 of table 3 are fairly low for this type of exercise, about 40% instead of about 80% in Duranton *et al.* (2014) for the US. This is mostly caused by sampling errors. The number of pairwise trade links is around 7,700 involving more than 900 municipalities. For the US, Duranton *et al.* (2014) rely on a much larger survey and far fewer geographic units. This said, sampling errors are not an issue per se. They will only lead to noisy estimates of municipal fixed effect, which are used as dependent variable in the second step. However, one might worry that

Table 4: Second-step results, OLS for exporter fixed effects, trade in weight

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Road index	0.30[a]	0.18[a]	0.21[a]	0.18[a]	0.21[a]	0.18[a]	0.20[a]	0.18[a]
	(0.050)	(0.052)	(0.055)	(0.058)	(0.063)	(0.065)	(0.060)	(0.067)
log 2010 population		0.51[a]	0.49[a]	0.49[a]	0.92[c]	0.80	1.07[b]	0.93[c]
		(0.098)	(0.10)	(0.10)	(0.54)	(0.61)	(0.43)	(0.54)
Market access (export)		-0.063	0.022	-0.16	0.23	0.020	0.091	-0.048
		(0.20)	(0.25)	(0.28)	(0.27)	(0.33)	(0.24)	(0.32)
log 2011 wage			0.65	0.88[c]	0.63	0.99		
			(0.51)	(0.52)	(0.70)	(0.74)		
log 1951 population					0.13	0.11	0.084	0.035
					(0.22)	(0.22)	(0.21)	(0.22)
log 1964 population					-0.16	-0.13	-0.18	-0.18
					(0.28)	(0.30)	(0.29)	(0.30)
log 1985 population					0.23	0.71	0.078	0.68
					(0.94)	(1.06)	(0.96)	(1.07)
log 1993 population					-0.53	-0.93	-0.47	-0.89
					(0.70)	(0.74)	(0.71)	(0.74)
Andean				-0.31		-0.31		-0.23
				(0.27)		(0.26)		(0.30)
Carribean				0.12		0.22		0.030
				(0.35)		(0.40)		(0.35)
Oriental				0.43		0.44		0.74
				(0.31)		(0.45)		(0.53)
R^2	0.18	0.34	0.39	0.42	0.41	0.45	0.41	0.45
Observations	134	134	119	115	96	92	102	98

Notes: All regressions include a constant. Robust standard errors in parentheses unless otherwise mentioned. [a], [b], [c]: significant at 1%, 5%, 10%.

trade with smaller cities is less frequently reported because of the placing of weigh stations. To minimize this risk, second-step regressions are estimated for cities with population above 40,000. This represents a compromise between the necessity to keep a large-enough sample of cities and getting reliable estimates of trade flows. The robustness of the main results with respect to other samples of cities is assessed below.

5. The effects of road on trade: second-step results

5.1 OLS *results*

I now turn to the estimation of equation (21) and its counterpart for trade in weight to assess the effects of roads within cities on their propensity to export. Table 4 reports results for OLS regressions using the exporter fixed effects estimated in the first step for trade in weight as dependent variable. More specifically, this table uses the fixed effects for trade in weight estimated from column 1 of panel A of table 3 for municipalities with population above 40,000. The key explanatory variable of interest is a road index that sums the log of the number of principal roads, the log of the mileage of these roads, and the log of the number of exists from the municipality by principal roads. It is useful to think of the coefficient on this variable as an elasticity of trade with respect to city roads in a broad sense. I verify below the robustness of the results for alternative first-step specifications, alternative samples of municipalities, and alternative measures of roads.

Column 1 is a rudimentary specification using only the road index as explanatory variable. Column 2 adds population and market access. Market access is computed as suggested by equation (12).[15] Treating productivity as a missing variable, this specification can be viewed as the empirical implementation of the equivalent of equation (21) for trade in weight, itself the direct empirical counterpart of equation (18) in the model. Column 3 also include wages. Wages are measured as the municipal fixed effect of a wage equation that corrects for individual characteristics in 2011 as in Duranton (2013). This variable may be thought of as an imperfect proxy for productivity.[16] Column 4 further includes 3 regional indicator variables for the Andean, the Carribean, and the Oriental regions of Colombia (with the Pacific region being the reference). Column 5 adds instead population from 1951, 1964, 1985, and 1993 to proxy implicitly for unobserved city effects that may also determine roads. Column 6 considers both regional indicators and past population. This is the most demanding specification. Finally, columns 7 and 8 repeat columns 5 and 6 without the wage variable.

The coefficient on roads is highly significant in all columns and fluctuates between 0.18 and 0.30. Ignoring the first rudimentary specification of column 1, this coefficient is extremely stable

[15] The computation of market access uses the results from the same first-step specification. A city's own importer effect is not included in this computation to avoid endogeneity biases.

[16] The wage variable has an ambiguous status here. If it is treated as the wage in the model, it actually calls for a different second-step regression where neither roads nor market access are included. Alternatively, one could think of our measure of wages as a proxy for productivity.

and only fluctuates between 0.18 and 0.21 in columns 2-8. In particular, it is equal to 0.18 in column 2 which is the specification most closely connected to the model. It is also equal to 0.21 in column 3 which adds wages and to 0.18 again in column 6, the most complete specification.

A number of other results are also of interest. First, equation (18) interprets the coefficient on population as $\frac{\sigma}{\sigma+\gamma}$. In columns 2 to 7 where the coefficient on population is estimated, it fluctuates from about 0.5 in columns 2-4 to 1.07 in column 7. Because population is serially correlated, it is unsurprising that the coefficient on population in columns 5 to 8 is either weakly significant or insignificant because of past population. Without the controls for past populations in columns 2 to 4, the coefficient is about 0.5 and highly significant. This would suggest that elasticity of substitution between goods, σ is equal to the labour supply elasticity in the formal sector, γ. Equation (8) indicates that the coefficient on distance in the first step should be interpreted as $1-\sigma$, which suggests a value of 1.6 to 1.7 for σ and thus for γ.[17] Equation (18) also predicts that the coefficient on wages (interpreted as productivity) should be equal to $1+\gamma$ times the coefficient on population. The wage coefficient in column 3 to 6 varies between 0.63 and 0.99 but it is significant only once. These point estimates suggest a value of γ from 1.2 to 1.8, which is consistent with the coefficient on population. Finally, equation (18) predicts the same coefficient for market access as for wages but opposite in sign. In table 4, the coefficient on market access is negative or close to zero but generally insignificant. Given the large standard errors around this coefficient, we cannot reject the prediction of equation (18). These more structural interpretations of our results should be, of course, taken with a grain of salt. They nonetheless fail to reveal any major discrepancy between the model and the results.

In results not reported here, I also experimented with a variety of other city characteristics including geography. As with the analysis of wages in Duranton (2013), the results for those variables are generally disappointing and do not affect the coefficient on roads. The only mildly robust association with trade can be found for altitude. Cities that are located neither too high nor too low appear to trade more. This association between trade and altitute nonetheless disappears in the instrumental variable regressions of the type estimated below.

It is also interesting to compare the results here with those of Duranton *et al.* (2014) who rely

[17]Broda and Weinstein (2006) estimate a wide range of coefficients for σ with a large majority of them being between 2 and 10. This suggests a considerably higher, but not implausible, value for γ. As argued above, there are good reasons for a lower estimate of the elasticity of substitution in Colombia. Only goods with a low σ may be traded. The lack of competition may also lead to high producer margins that are reflected in low estimates for σ.

Table 5: Second-step results, OLS for exporter fixed effects, trade in value

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Road index	0.33^a	0.22^a	0.25^a	0.22^a	0.25^a	0.22^a	0.25^a	0.22^a
	(0.055)	(0.060)	(0.064)	(0.071)	(0.070)	(0.076)	(0.066)	(0.076)
log 2010 population		0.47^a	0.46^a	0.46^a	0.98	0.80	1.05^b	0.83
		(0.10)	(0.10)	(0.11)	(0.61)	(0.66)	(0.50)	(0.58)
Market access (export)		-0.16	-0.13	-0.35	0.066	-0.20	-0.039	-0.24
		(0.21)	(0.25)	(0.28)	(0.30)	(0.35)	(0.27)	(0.34)
log 2011 wage			0.51	0.62	0.49	0.62		
			(0.54)	(0.56)	(0.77)	(0.82)		
log 1951 population					0.14	0.11	0.088	0.037
					(0.21)	(0.23)	(0.21)	(0.23)
log 1964 population					-0.20	-0.19	-0.20	-0.20
					(0.25)	(0.26)	(0.26)	(0.26)
log 1985 population					0.27	0.87	0.18	0.93
					(0.99)	(1.14)	(0.96)	(1.10)
log 1993 population					-0.63	-1.05	-0.58	-1.06
					(0.85)	(0.91)	(0.80)	(0.86)
Andean				-0.050		-0.0013		0.026
				(0.31)		(0.30)		(0.33)
Carribean				0.33		0.42		0.25
				(0.37)		(0.42)		(0.36)
Oriental				0.63		0.81		1.00^c
				(0.41)		(0.54)		(0.59)
R^2	0.20	0.32	0.36	0.39	0.40	0.43	0.39	0.42
Observations	134	134	119	115	96	92	102	98

Notes: All regressions include a constant. Robust standard errors in parentheses unless otherwise mentioned. ^{a}, ^{b}, ^{c}: significant at 1%, 5%, 10%.

on the same methodology. The main coefficient of interest for the road index is lower here: 0.18 in column 7 instead of 0.38 for a similar specification in Duranton *et al.* (2014). Keeping in mind that the road index is computed slightly differently in Colombia and the possible endogeneity of this variable, a more complete discussion of this difference is postponed until later. The behaviour of the coefficient on population is very similar. Finally, Duranton *et al.* (2014) estimate negative coefficients for market access which are much larger in absolute value than those obtained here and highly significant. This difference could be due to market access being poorly measured in Colombia when using the output of the first-step gravity regression.

Table 5 duplicates the regressions of table 4 but uses fixed effects for trade in value estimated

Table 6: Second-step results, OLS for importer effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Trade in weight.								
Road index	0.33[a]	0.20[a]	0.19[a]	0.16[b]	0.21[a]	0.18[a]	0.23[a]	0.18[a]
	(0.054)	(0.058)	(0.062)	(0.066)	(0.062)	(0.067)	(0.060)	(0.062)
Controls								
2010 population	N	Y	Y	Y	Y	Y	Y	Y
Market access	N	Y	Y	Y	Y	Y	Y	Y
2011 wage	N	N	Y	Y	Y	Y	N	N
Past populations	N	N	N	N	Y	Y	Y	Y
Region indicators	N	N	N	Y	N	Y	N	Y
R^2	0.21	0.35	0.39	0.42	0.46	0.52	0.45	0.53
Observations	134	134	119	115	96	92	102	98
Panel B. Trade in value.								
Road index	0.26[a]	0.14[b]	0.19[a]	0.16[a]	0.22[a]	0.19[a]	0.24[a]	0.18[a]
	(0.060)	(0.061)	(0.059)	(0.059)	(0.065)	(0.066)	(0.059)	(0.058)
R^2	0.22	0.36	0.40	0.44	0.46	0.52	0.45	0.52
Observations	134	134	119	115	96	92	102	98

Notes: All regressions include a constant. Robust standard errors in parentheses unless otherwise mentioned. [a], [b], [c]: significant at 1%, 5%, 10%.

from column 1 of panel B of table 3 instead of fixed effects for trade in weight. The results regarding the road index are very similar to those of table 4. For all the main explanatory variables, the coefficients of table 5 are all within a standard deviation of their corresponding coefficient in table 4. Interestingly, the coefficient on roads is always higher by a small amount in table 5. This difference is nonetheless tiny and the broad similarity between the coefficients for trade in weight and trade in value strongly suggests that roads within a municipality do not affect patterns of specialisation. This similarity of results for trade in weight and trade in value is in contrast with the results of Duranton *et al.* (2014) where there is no apparent effect of roads on trade in value. Again, it is better to have a fuller overview of the results before commenting on this difference.

It is also useful to compare results for exports with those for imports. Panels A and B of table 6 duplicate the regressions of tables 4 and 5 using importer instead of exporter fixed effects. Despite the correlations between exporter and importer effects being 'only' about 0.5 to 0.6, the road coefficients are remarkably similar. The match is nearly perfect for trade in weight. The estimated coefficients are slightly lower for imports in value relative to those for exports but they are usually

Table 7: Second-step results, LIML for exporter effects, trade in weight

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Instruments from the 1938 road network.								
Road index	0.55[a]	0.49[c]	0.33[b]	0.23	0.25	0.16	0.23[c]	0.11
	(0.15)	(0.25)	(0.13)	(0.17)	(0.17)	(0.21)	(0.14)	(0.16)
Controls								
2010 population	N	Y	Y	Y	Y	Y	Y	Y
Market access	N	Y	Y	Y	Y	Y	Y	Y
2011 wage	N	N	Y	Y	Y	Y	N	N
Past populations	N	N	N	N	Y	Y	Y	Y
Region indicators	N	N	N	Y	N	Y	N	Y
Overid. p-value	0.86	0.19	0.15	0.14	0.13	0.10	0.27	0.26
First-stage stat.	17.8	6.45	11.2	6.17	13.8	7.16	14.9	8.10
Observations	134	134	119	115	96	92	102	98
Panel B. Instruments from caminos reales.								
Road index	0.43[b]	0.29	0.32[b]	0.17	0.22	0.086	0.20	0.0056
	(0.17)	(0.20)	(0.14)	(0.19)	(0.19)	(0.24)	(0.17)	(0.19)
Overid. p-value	0.69	0.34	0.50	0.87	0.35	0.58	0.43	1.00
First-stage stat.	8.13	6.42	11.8	6.25	12.9	7.46	12.3	7.83
Observations	134	134	119	115	96	92	102	98

Notes: All regressions include a constant. Robust standard errors in parentheses unless otherwise mentioned. [a], [b], [c]: significant at 1%, 5%, 10%. Instruments are the number of kilometers of roads and distance to the nearest road for 1938 roads (in panel A) or caminos reales (in panel B).

not statistically different.

5.2 IV *results*

Table 7 mirrors table 4 but uses LIML estimation and instruments the index for contemporaneous city roads with, respectively, instruments from 1938 road network in panel A and instruments from the map of caminos reales in panel B.[18] Namely, table 7 uses the number of kilometers of historical roads and the distance to the nearest roads. For municipalities served by historical roads, this second instrument is zero whereas, for municipalities without historical roads, it is the first instrument which is equal to zero.

[18]Limited information maximum likelihood (LIML) is a one-stage IV estimator. Compared to TSLS, it provides more reliable point estimates and test statistics with weak instruments. As for estimating instrumental variable regressions with the generalized method of moments, it has a clear advantage in the presence of heteroskedasticity of unknown form.

Depending on the specification, the coefficient on the contemporaneous road network is larger or lower than the corresponding OLS coefficients reported in table 4. In the first three columns of panel A, which uses instruments from the 1938 road network, the coefficient on roads is larger than under OLS and significant. It is equal to 0.49 in column 2 and 0.33 in column 3. In columns 4-6 and 8, this coefficient becomes insignificant but the point estimates are close to the corresponding OLS estimates. In column 7, the coefficient on roads is weakly significant and marginally higher than its OLS counterpart. In panel B, which uses instruments from the colonial road network, the results are very similar but with marginally less significance overall.

This is a mixed picture for trade in weight. These IV estimations are consistent with the OLS results but the lack of strength of the instruments leads to more noisy estimates which are sometimes not statistically different from zero and usually not statistically different from their OLS counterparts. As reported in the table, the instruments are strong in about half the cases for 1938 roads and slightly less often for caminos reales and marginally weak in the other cases.

Table 8 duplicates table 7 but uses fixed effects from the first step for trade in value as opposed to trade in weight. In panel A, the instruments remain the number of kilometers of historical roads and distance to the nearest roads for the 1938 network whereas panel B reports results using the same instruments computed from the map of caminos reales of the colonial era. In panel A, the coefficient on roads is significant in six of the columns and ranges from 0.35 to 0.59. In panel B with colonial roads instruments, this coefficient is again significant for six specifications in eight. The significant coefficients range from 0.35 to 0.46. Overall, the coefficients on the road index are higher with LIML than with OLS. Among the LIML estimations, 1938 instruments yield on average slightly larger coefficients although the differences between the instruments are not significant. Just like with OLS, the coefficients on roads are larger for trade in value than for trade in weight and instrumenting appears to have magnified these differences slightly.[19]

Although we do not report these coefficients here to keep the tables easy to read, the coefficients on the other explanatory variables in tables 7 and 8 are very similar to those of tables 4 and 5.

Overall, the IV results of tables 7 and 8 broadly confirm the OLS results reported in tables 4 and 5 and are suggestive of perhaps larger roads effects on trade, especially for values. Although these differences should not be overstated since they are in many cases statistically insignificant,

[19]Because LIML is a one-step estimation procedure, the 'first stage' is not exactly the same regardless of the dependent variable as it would be with TSLS.

Table 8: Second-step results, LIML for exporter effects, trade in value

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Instruments from the 1938 road network.								
Road index	0.59^{a}	0.56^{b}	0.49^{a}	0.44^{b}	0.46^{b}	0.44	0.42^{a}	0.35
	(0.15)	(0.24)	(0.16)	(0.22)	(0.22)	(0.32)	(0.16)	(0.22)
Controls								
2010 population	N	Y	Y	Y	Y	Y	Y	Y
Market access	N	Y	Y	Y	Y	Y	Y	Y
2011 wage	N	N	Y	Y	Y	Y	N	N
Past populations	N	N	N	N	Y	Y	Y	Y
Region indicators	N	N	N	Y	N	Y	N	Y
Overid. p-value	0.90	0.18	0.10	0.069	0.032	0.024	0.088	0.058
First-stage stat.	17.8	6.55	11.2	5.99	13.5	6.93	14.5	7.69
Observations	134	134	119	115	96	92	102	98
Panel B. Instruments from caminos reales.								
Road index	0.46^{a}	0.35^{c}	0.46^{a}	0.39^{b}	0.44^{b}	0.37	0.40^{b}	0.28
	(0.16)	(0.20)	(0.15)	(0.19)	(0.22)	(0.27)	(0.18)	(0.19)
Overid. p-value	0.78	0.33	0.40	0.64	0.21	0.31	0.29	0.60
First-stage stat.	8.13	6.44	11.7	6.08	12.7	7.22	12.0	7.42
Observations	134	134	119	115	96	92	102	98

Notes: All regressions include a constant. Robust standard errors in parentheses unless otherwise mentioned. ^{a}, ^{b}, ^{c}: significant at 1%, 5%, 10%. Instruments are the number of kilometers of roads and distance to the nearest road for 1938 roads (in panel A) or caminos reales (in panel B).

they could be indicative of two things. First, our instruments might be correcting for classical measurement error. While the data that underlie the map of contemporaneous roads of figure 2 are arguably accurate, the digitisation of this map may have introduced mild measurement errors, due in part to the arbitrary and sometimes capricious nature of municipal borders in Colombia.[20] That the road index may not accurately describe the transport situation of Colombian municipalities is a more serious form of measurement error that the instruments may also correct.

Roads may have been allocated to places that trade less. While caminos reales coincide with major contemporary trade flows between the main cities of the interior and both coasts and the 1938 network mainly represents roads between neighbouring cities of importance, more recent expansions of the network may have been more 'political' in nature. For instance, the current road

[20]In some cases, roads go along municipal borders and end up crossing them several times which affects the 'exit' variable used in the construction of the road index. The data was kept as originally processed to avoid arbitrary adjustments.

network contains the relatively recent road between the two largest cities of the country, Bogotà and Medellìn, which was absent from the map of caminos reales and from the 1938 road network. These two cities do not trade much with each other given their size and the distance that separates them (Roda *et al.*, 2012). Other recent extensions of the road network in Colombia have been towards isolated areas of the Northwest of the country, perhaps to help them integrate more. This suggest a mild bias towards the allocation of roads to places that trade little.

To complete these IV results, panel A of table 18 duplicates the specification of panel A of table 7 for importer fixed effects instead of exporter fixed effects. The results are marginally weaker. The coefficients on the road index are on average slightly lower and less significant. Panel B of table 18 duplicates again the same set of regressions but uses importer fixed effects for trade in value. Relative to panel A of table 8, the effects are also weaker.

These results for Colombia can now be compared to the results of Duranton *et al.* (2014) for the US. Subject to minor differences in variable definitions, the coefficient on roads of 0.49 from column 2 of table 7 for trade in weight is directly comparable to a coefficient of 0.57 for the US while the coefficient of 0.23 in column 7 is directly comparable to a coefficient of 0.47 for the US. These two coefficient are thus somewhat lower in Colombia than in the US. Turning to trade in value, the difference is in the opposite direction. The coefficient on roads of 0.56 from column 2 of table 8 for trade in weight is directly comparable to an insignificant coefficient of 0.17 for the US while the coefficient of 0.42 in column 7 is directly comparable to an insignificant coefficient of 0.07 for the US.

The main difficulty with this comparison is that trade within Colombia takes place over distances that are typically shorter than in the US, at most about 1,000 kilometres in the data used here. For instance, the distance between Barranquilla, the main port on the Atlantic coast, and Buenaventura, the main port on the Pacific coast, is about 900 kilometres. This corresponds to the distance between Las Vegas and San Francisco. When they conduct their estimations for trade taking place over less than 1,000 kilometres, Duranton *et al.* (2014) find a coefficient of 0.75 for weight and 0.29 for value in the US. While this increases the gap with my coefficient for trade in weight in Colombia, this makes the results for trade in value much more similar. Regardless of the range of distances that are considered, roads affect the specialisation of US cities but not that of

Colombian cities.[21]

Linking back to the theoretical model above, recall that section 4 already provided strong support for the prediction that a reduction in road distance between cities leads to an increase in trade between them. The results of this section provide support for the third prediction of the model that an increase in roads within a city leads to an increase in the weight and value of trade. As argued by the second prediction of the model, this increase in export comes about through an increase in the demand for labour which leads to higher wages and an expansion of employment in the formal sector. In absence of an expansion of employment, the increase in the value of trade predicted by the model should be equal to the increase in wages since export revenues are eventually dissipated in wages. An elasticity of wages with respect to roads of about 0.5 would be extremely large. Elsewhere (Duranton, 2013), I explore the effect of roads on wages but fail to uncover elasticities larger than 0.05. This suggests small effects of roads on wages. For roads to lead to large increases in exported values, the extensive margin of adjustment working through increased employment in formal sectors must be powerful. This is not inconsistent with the interpretations of the coefficients on wages given above suggesting a potentially large value of γ.

The effect of within-city highways on the weight and value of exports is economically large. To assess their magnitude more precisely consider a city, with the mean road index of 3.97, and another city, one standard deviation above with a road index of 5.50. Given the logs, the difference between these two cities corresponds to a factor of 4.5. Because they are closest to our theoretical specification, it is best to focus on the results from columns 2 or 3 of each table. Using the OLS coefficient of 0.18 in column 2 of table 4, this one standard deviation in the road index implies an increase in the weight of exports of 32%. From column 2 of table 5, this standard deviation increase in the road index also implies an increase in the value of exports of 40%. Using the IV coefficient of column 2 of panel A of tables 7 and 8, the estimated values of 0.49 and 0.56 imply increases in the weight and value of exports of 112% and 136%, respectively.

Note finally that from the estimates of table 3, the elasticity of trade in weight with respect to driving time is -0.62. The corresponding elasticity for trade in value is -0.69. That is, a halving of driving time between two cities increases trade between them by 53% in weight and 61% in value.

[21] In addition, Duranton *et al.* (2014) find no effect of city roads on the weight and value of imports, consistently with their main specialisation results. For Colombia, there is weak evidence of roads causing an increase in imports of similar magnitude for both weight and value. This is the natural counterpart to increased exports.

A doubling of within-city roads increases trade in weight and value by 13 and 16% respectively according to the OLS estimates and by 40 and 47% according to the LIML estimates. Hence, in many ways, the results for within city roads are not out of line with those for roads between cities. Roads both between and within cities seem to matter a lot.

5.3 *Robustness checks*

I now verify the robustness of the main second-step results along a number of dimensions.

As with all instrumental variables estimations, there are some worries with the instruments. As discussed above, the instruments are not overly strong. While first-stage statistics of 6 or 7 would be considered weak with two-stage least squares, they are still acceptable with LIML, an estimation procedure which is more robust to weak instruments.[22] Recall also that the results are very similar when the instruments are strong and have a first-stage statistic around 14. It is also the case that two different sets of instruments from the networks of colonial roads and from the 1938 road network yield very similar results. As argued above and as can be visually detected from figure 2, these two networks are very different, one appears to be a good predictor of long distance links between major cities (colonial roads) while the other is a good predictor of links between cities within the same region (1938 roads). Despite these differences, the results from both sets of instruments are very close. Instruments from the 1938 road network tend to yield significant coefficients more often but the insignificant point estimates obtained from colonial roads are usually similar.

Note also that once current population is added as a control, the instrumented coefficient on roads does not change much with the inclusion of further controls including wages, past population, and regional dummies. The coefficient on roads sometimes becomes insignificant but this appears to be caused by decreasing sample size as further control variables are added. This stability of the instrumented road coefficient when geographical, population, and economic variables are added is reassuring. This suggests that any missing variable potentially correlated with roads is not strongly correlated with any of our controls. This considerably reduces the scope for our instruments to be invalid. Put differently, for the strategy proposed here to go wrong, we

[22]For instance, the 10% critical value of the size test for two instruments is 8.68 and the 15% critical value s 5.33 according to Stock and Yogo (2005) .

Table 9: Second-step results, LIML for exporter fixed effects with various combinations of instruments, trade in value

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Road index	0.74^{b}	0.55^{c}	0.65^{b}	0.64^{a}	0.48^{b}	0.36^{b}	0.82^{c}	0.69^{b}
	(0.34)	(0.31)	(0.26)	(0.22)	(0.24)	(0.16)	(0.43)	(0.33)
Instruments Distance	Y	log		Y				Y
Length		Y	/area	Y	Y		log	/area
Frequency				Y	log		log	Y
Exits						Y	log	Y
Overid. p-value	-	0.35	-	0.37	0.10	0.27	0.12	0.14
First-stage stat.	7.75	6.38	11.0	5.97	5.85	4.03	2.98	4.39
Observations	134	68	130	134	66	134	65	130

Notes: All regressions include a constant. Robust standard errors in parentheses unless otherwise mentioned. [a], [b], [c]: significant at 1%, 5%, 10%. Instruments from the 1938 road network as described in the text.

would need a missing variable that is correlated in the same way with both sets of instruments and uncorrelated with the control variables. While possible, this seems unlikely.

To explore this issue more in-depth, table 9 examines the robustness of the coefficient on roads in column 2 of table 8 for trade in value by varying the set of instruments. The coefficient on the road index is equal to 0.56 in the reference specification of table 9. Column 1 of table 9 uses only distance to 1938 roads as an instrument and yields a slightly higher coefficient of 0.74. Column 2 uses log distance and the number of kilometers of 1938 roads to obtain the same coefficient as when distance is expressed in level. While the 1938 road mileage is a weak instrument on its own, it can be normalised by the area of the municipality to become strong and yield a coefficient of 0.65 in column 3. Column 4 adds the number of different road segments in 1938 to distance and mileage. The coefficient on roads remains stable at 0.64. Using the number of 1938 road segments in logs with the 1938 mileage yields a coefficients of 0.48 in column 5. Column 6 uses the number of 1938 road exits alone. This instrument is marginally weak but still yields a significant coefficient of 0.36, which is still not statistically different from our reference of 0.56. Column 7 uses 1938 mileage, number of segments, and exits in log to estimate a coefficient on roads somewhat higher at 0.82. Finally, column 8 uses 4 instruments from 1938 and retains a coefficient of 0.69. While there is some variability between those estimates, they are all of similar magnitude and statistically within one standard error of each other.

Table 10: Second-step results, OLS for exporter effects in large cities

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Trade in weight.								
Road index	0.29^a	0.20^a	0.19^b	0.15^c	0.082	0.060	0.12	0.040
	(0.062)	(0.070)	(0.077)	(0.079)	(0.088)	(0.091)	(0.079)	(0.088)
Controls								
2010 population	N	Y	Y	Y	Y	Y	Y	Y
Market access	N	Y	Y	Y	Y	Y	Y	Y
2011 wage	N	N	Y	Y	Y	Y	N	N
Past populations	N	N	N	N	Y	Y	Y	Y
Region indicators	N	N	N	Y	N	Y	N	Y
R^2	0.24	0.35	0.37	0.50	0.44	0.55	0.35	0.51
Observations	56	56	55	53	46	44	47	45
Panel B. Trade in value.								
Road index	0.31^a	0.24^a	0.23^a	0.20^b	0.15^c	0.14	0.18^b	0.12
	(0.074)	(0.080)	(0.085)	(0.092)	(0.086)	(0.094)	(0.081)	(0.089)
R^2	0.25	0.35	0.37	0.49	0.39	0.49	0.34	0.47
Observations	56	56	55	53	46	44	47	45

Notes: Cities with population above 100,000 only. All regressions include a constant. Robust standard errors in parentheses unless otherwise mentioned. ^{a}, ^{b}, ^{c}: significant at 1%, 5%, 10%.

A second possible worry pertains to the sample of cities. The analysis above focuses on cities with population above 40,000 endowed with roads in 2010. Panel A of table 10 duplicates the regressions of table 4 but considers only cities with population above 100,000.[23] In the first four specifications, the results are very close across the two tables. The same patterns holds for trade in value for which the results are reported in panel B and can be compared to those of table 5. For columns 5 to 8, for which the specifications are more demanding, the coefficient on roads become insignificant. This is possibly caused by a lack of statistical power since the number of observations is lower for these specifications. Column 7, for instance, uses only 44 observations and includes 11 explanatory variables.

Table 11 repeats the same exercise but only considers smaller cities with population between 20,000 and 50,000. For the first two columns, the estimated OLS coefficient on roads is lower than in tables 4 and 5. The results in columns 3 to 8 are more similar with the caveat that for trade in value the larger standard errors make them insignificant. An issue in this table is that the control

[23]Table 10 just like table 11 below only reports OLS results. The sample sizes get much smaller. Then, the IV results get noisy and comparisons with main results become uninformative.

Table 11: Second-step results, OLS for exporter effects in small cities

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Trade in weight.								
Road index	0.11^c	0.11^c	0.21^b	0.19^c	0.24^b	0.23^c	0.15^c	0.17^c
	(0.064)	(0.064)	(0.10)	(0.10)	(0.12)	(0.12)	(0.088)	(0.093)
Controls								
2010 population	N	Y	Y	Y	Y	Y	Y	Y
Market access	N	Y	Y	Y	Y	Y	Y	Y
2011 wage	N	N	Y	Y	Y	Y	N	N
Past populations	N	N	N	N	Y	Y	Y	Y
Region indicators	N	N	N	Y	N	Y	N	Y
R^2	0.01	0.07	0.14	0.20	0.27	0.29	0.14	0.14
Observations	168	168	82	81	60	59	114	113
Panel B. Trade in value.								
Road index	0.13^c	0.13^c	0.20	0.21	0.17	0.16	0.13	0.15
	(0.071)	(0.070)	(0.12)	(0.13)	(0.15)	(0.15)	(0.10)	(0.11)
R^2	0.02	0.04	0.09	0.14	0.13	0.17	0.04	0.06
Observations	168	168	82	81	60	59	114	113

Notes: Cities with population between 20,000 and 50,000 only. All regressions include a constant. Robust standard errors in parentheses unless otherwise mentioned. ^{a}, ^{b}, ^{c}: significant at 1%, 5%, 10%.

variables used in columns 3 to 8 are not available for many small cities so that sample size varies considerably across specifications. These cities, for which wages and past populations are not available, tend to be more peripheral. It is possible that more peripheral cities benefit less from road access.

To check this conjecture further, panel A of table 12 duplicates the regressions for trade in weight of table 7 but only considers trade flows above 100 kilometres in the first step to estimate the propensity of cities to export. Panel B of table 12 performs a similar exercise for a higher censoring threshold for trade flows of 200 kilometres. The results are interesting. For trade flows above 100 kilometers in panel A, the results are very similar to those considering all trade flows in table 8. For trade flows above 200 kilometres, the coefficients on city roads gets lower and the results are overall weaker. This suggests that city roads matter more for trade with nearby cities than for long distance trade.

Although it is at odds with the multiplicative assumptions regarding the effects of transport costs made for the model above to remain tractable, this result is intuitive. The iceberg nature of

Table 12: Second-step results, LIML for exporter effects with distance censoring

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Trade in value, 100 kilometre censoring, and instruments from the 1938 road network.								
Road index	0.53^a	0.41^c	0.35^b	0.29	0.28^c	0.24	0.30^b	0.26
	(0.13)	(0.23)	(0.15)	(0.18)	(0.17)	(0.19)	(0.14)	(0.16)
Controls								
2010 population	N	Y	Y	Y	Y	Y	Y	Y
Market access	N	Y	Y	Y	Y	Y	Y	Y
2011 wage	N	N	Y	Y	Y	Y	N	N
Past populations	N	N	N	N	Y	Y	Y	Y
Region indicators	N	N	N	Y	N	Y	N	Y
Overid. p-value	0.41	0.60	0.45	0.38	0.13	0.094	0.30	0.23
First-stage stat.	16.9	5.36	9.32	5.60	9.50	6.63	11.3	7.78
Observations	128	128	114	111	92	89	98	95
Panel B. Trade in value, 200 kilometre censoring, and instruments from the 1938 road network.								
Road index	0.50^a	0.060	0.21	0.15	-0.028	-0.14	-0.049	-0.16
	(0.19)	(0.25)	(0.19)	(0.29)	(0.23)	(0.29)	(0.23)	(0.31)
Overid. p-value	0.0089	0.37	0.74	0.76	0.92	0.90	0.78	0.71
First-stage stat.	14.9	4.55	9.05	4.54	8.62	5.34	8.94	5.32
Observations	121	121	108	106	86	84	91	89

Notes: All regressions include a constant. Robust standard errors in parentheses unless otherwise mentioned. ^{a}, ^{b}, ^{c}: significant at 1%, 5%, 10%. Instruments are the number of kilometers of 1938 roads and distance to the nearest 1938 road.

transport costs assumed above implies that the cost of exiting a city should be relatively higher for shipments over longer distances.[24] This is of course questionable and the results of table 12 should be read as supportive of the notion that transport costs may not be fully multiplicative in nature. This finding is consistent with the results of Duranton *et al.* (2014) for trade between US cities. They show that city roads have a large effects on shipments to nearby cities but no statistically discernible effect on shipments to remote cities.

Another series of worries relates to the main explanatory variable. To estimate an elasticity, we need to use logs values for the number of kilometers of roads, the number of roads, and the number of exits out of the municipality. A small number of municipalities do not have roads and

[24]If for instance, 10% of the value of a shipment is lost to exit a city, this 10% will represent a much bigger proportion of what will be left of this shipment when it arrives at a faraway destination than a close destination. For instance for a shipment of a value 100 at origin, the cost of exiting may represent 10 for a value of 40 upon arrival at a remote destination instead of 10 of 80 for a close destination.

Table 13: Second-step results, OLS for exporter effects using an alternative index and including municipalities with no roads

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Trade in weight.								
Road index′	0.22^a	0.11^c	0.18^a	0.13^b	0.19^b	0.15^b	0.16^b	0.12^c
	(0.055)	(0.055)	(0.062)	(0.064)	(0.074)	(0.073)	(0.067)	(0.067)
Controls								
2010 population	N	Y	Y	Y	Y	Y	Y	Y
Market access	N	Y	Y	Y	Y	Y	Y	Y
2011 wage	N	N	Y	Y	Y	Y	N	N
Past populations	N	N	N	N	Y	Y	Y	Y
Region indicators	N	N	N	Y	N	Y	N	Y
R^2	0.13	0.31	0.37	0.41	0.40	0.45	0.39	0.44
Observations	141	141	123	119	99	95	106	102
Panel B. Trade in value.								
Road index′	0.24^a	0.13^c	0.23^a	0.18^b	0.24^b	0.19^b	0.20^b	0.16^c
	(0.068)	(0.069)	(0.075)	(0.080)	(0.091)	(0.095)	(0.085)	(0.088)
R^2	0.13	0.28	0.36	0.39	0.39	0.43	0.37	0.41
Observations	141	141	123	119	99	95	106	102

Notes: All regressions include a constant. Robust standard errors in parentheses unless otherwise mentioned. ^{a}, ^{b}, ^{c}: significant at 1%, 5%, 10%. Road index′ sums log 1+ number of exits , log 1+ number of roads, and log 1+ number of kilometres roads.

thus have been excluded from the estimation. We can include them by considering an alternative road index that sums that sums the log of 1 + the number of principal roads, the log of 1+ the mileage of roads, and the log of 1+ the number of exist from the municipality by principal roads. This alternative transformation of the components of the road index also provides a robustness check regarding the functional form that was chosen above.

The results are reported in table 13. Panel A reports OLS results for trade in weight which mirror those of table 4 while panel B reports OLS results for trade in value which mirror those of table 5. The coefficient on this alternative road index varies between 0.11 and 0.22 in panel A and between 0.13 and 0.24 in panel B. This is below the corresponding coefficients of tables 4-5.

However, the standard deviation for the road index used in table 13 is 2.50 against only 1.53 for the road index used in tables 4 and 5. Hence, a standard deviation in the road index is associated 0.45 log point in the propensity to export weight in column 2 of table 13. For the same specification in table 4, the effect of a standard deviation in the road index is 0.50 log point in the propensity

Table 14: Second-step results, LIML for exporter effects with distance censoring

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Trade in weight, alternative first-step specifications								
Road index	0.49^{c}	0.41^{c}	0.41^{c}	0.44^{c}	0.46^{b}	0.42^{c}	0.47^{c}	0.44^{c}
	(0.25)	(0.22)	(0.22)	(0.23)	(0.23)	(0.22)	(0.25)	(0.23)
Overid. p-value	0.19	0.28	0.29	0.23	0.23	0.26	0.22	0.27
First-stage stat.	6.45	6.37	6.32	6.61	6.62	6.39	6.46	6.58
Observations	134	134	134	134	134	134	133	133
Panel B. Trade in value, alternative first-step specifications								
Road index	0.56^{b}	0.47^{b}	0.46^{b}	0.51^{b}	0.52^{b}	0.48^{b}	0.54^{b}	0.51^{b}
	(0.24)	(0.21)	(0.21)	(0.22)	(0.23)	(0.21)	(0.24)	(0.23)
Overid. p-value	0.18	0.28	0.30	0.22	0.21	0.27	0.21	0.23
First-stage stat.	6.55	6.46	6.41	6.63	6.62	6.48	6.54	6.45
Observations	134	134	134	134	134	134	133	133

Notes: All regressions include a constant, 2010 population, and market access as controls duplicating the specification of column 2 of table 7 for panel A and table 8 in panel B. The dependent variable for each column is a series of fixed effects estimated in the corresponding column of table 3. Instruments are the number of kilometers of 1938 roads and distance to the nearest 1938 road.

to export weight. For the propensity to export value, a standard deviation in the road index is associated with 0.58 log point more trade in table 13 and 0.74 log point more trade in table 5. These numbers are fairly close and suggestive that neither the exact specification of the road index nor the sample of cities being studied matters, provided these cities are above a certain size.

The next series of worries concerns the dependent variable. Recall that second-step regressions use fixed effects estimated in the first step as dependent variables. To verify that our choice of a particular first-step specification does not matter, table 14 duplicates the specification of column 2 of table 7 in panel A for trade in weight and the equivalent specification from table 8 in panel B for trade in value for alternative first-step specifications. Specifically, each column of table 14 uses as dependent variable the fixed effects estimated in the corresponding column of table 3. As made clear by the table, the coefficients on the road index are extremely stable.

To provide further evidence on this issue, table 19 in appendix duplicates the specifications of table 7 in panel A and table 8 in panel B using the exporter fixed effects estimated in a first step that includes controls for product groups. More precisely, it uses the fixed effects estimated in the specification of column 1 of table 17 for trade in value and a corresponding specification for trade

Table 15: Second-step results, LIML for exporter effects, trade in value

	(1) weight OLS	(2) value OLS	(3) weight LIML	(4) value LIML	(5) weight OLS	(6) value OLS	(7) weight LIML	(8) value LIML
	Population market access				Income market access			
Road index	0.18^a	0.22^a	0.25^c	0.38^b	0.26^a	0.25^a	0.34^c	0.40^c
	(0.050)	(0.058)	(0.15)	(0.16)	(0.095)	(0.11)	(0.13)	(0.16)
Market access (export)	0.021	0.030	0.012	0.011	0.16^a	0.11	0.16^a	0.12
	(0.041)	(0.047)	(0.044)	(0.051)	(0.050)	(0.081)	(0.049)	(0.084)
R^2	0.39	0.37	-	-	0.63	0.51	-	-
Overid. p-value	-	-	0.40	0.33	-	-	0.38	0.33
First-stage stat.	-	-	8.01	8.01	-	-	5.05	5.05
Observations	124	124	124	124	62	62	62	62

Notes: All regressions include a constant and control for log 2010 population. Robust standard errors in parentheses unless otherwise mentioned. ^{a}, ^{b}, ^{c}: significant at 1%, 5%, 10%. Instruments are the number of kilometers of 1938 roads and distance to the nearest 1938 road.

in weight. Again, the results are very close to those of tables 7 and 8.

Finally, one may worry about the market access variable. In Duranton *et al.* (2014), it plays an important role and the road coefficient is sensitive to its inclusion. This does not appear to be the case here as the drop in the road index coefficient between column 1 and column 2 in tables 4-5 and 7-8 is not as large as in Duranton *et al.* (2014) and appears to be driven mostly by the inclusion of contemporaneous population. This said, table 15 explores these issues further by replicating the baseline specification of column 2 of tables 4-5 and 7-8 in columns 1-4 using an alternative measure of market access. For each city, this alternative measure sums the population of all other cities weighting them by inverse driving distance. A second alternative market access variable, which also weighs cities by their income, is used in columns 5 to 8.

As can be seen from the results of table 15, the coefficients on the road index are barely affected by these two alternative measures of market access. The coefficients on market access, on the other hand, differ. While the coefficient on the theory-based market access often negative (though insignificant), it is positive and significant for the income-based measure of market access for trade in weight.[25]

To sum-up, these robustness checks generally confirm the stability of our main OLS and IV estimates. The two novel findings are that within city roads seem to matter less for smaller cities

[25]See Duranton (2013) for a more an in-depth discussion of the differences between these different measure of market access.

and for longer distance trade.

6. Conclusions

This paper examines the causal effect of roads on the level and composition of trade in Colombian cities. I find that the weight and value of bilateral trade decrease rapidly with the highway distance between cities, with the weight of trade decreasing slightly less rapidly than its value. A 10% increase in the travel distance between trading partners reduces the value of trade between them by about 7% and the total weight by 6%. These large effects imply that reducing these pairwise distances through expanding the road system is likely to have large effects on trade and welfare.

We also find that within-city roads affect trade. A 10% increase in a city's stock of highways causes approximately a 3 to 5% increase in the weight and value of exports.

In sum, our results provide an important new insight into the way transport infrastructure affects trade and the organization of economic activity. Better transport encourages trade in general. Unlike in the US, it does not appear to affect specialisation. Instead, the key effects seem to be that the road infrastructure induces firms to trade between cities regardless of what they do.

While providing some answers, this paper also raised a number of issues to be considered by future research. The lower sensitivity of trade to distance in Colombia relative to the US is puzzling. This paper proposes two related conjectures about trade within Colombia being more 'Ricardian' than internal US trade and about less substitutable goods being traded in Colombia. Exploring these conjectures further would certainly be of interest. The model rationalises the fairly large effects of city roads on exports through a greater participation of the workforce in the formal sector where firms can export. Knowing more precisely which firms get to export more in cities with more roads would be important to understand the exact channels through which roads affect trade.

This paper is also focused on roads and internal trade. Colombia recently signed a free trade agreement with the US. Most long-distance international trade takes place by sea. Considering jointly internal and international trade is important task for future research to understand what the free trade agreement with the US will do to the location of economic activity in Colombia. Finally, this paper measures roads mostly in terms of length and ignores considerations of capacity and quality. These are likely to play an important role that future research should consider.

References

Alcalá, Francisco and Antonio Ciccone. 2004. Trade and productivity. *Quaterly Journal of Economics* 119(2):613–646.

Allen, Treb and Costas Arkolakis. 2014. Trade and the topography of the spatial economy. *Quarterly Journal of Economics* 129(3):forthcoming.

Anas, Alex, Richard Arnott, and Kenneth A. Small. 1998. Urban spatial structure. *Journal of Economic Literature* 36(3):1426–1464.

Anderson, James E. and Eric van Wincoop. 2003. Gravity with gravitas: A solution to the border puzzle. *American Economic Review* 93(1):170–192.

Arnott, Richard, André de Palma, and Robin Lindsey. 1990. Departure time and route choice for the morning commute. *Transportation Research Part B: Methodological* 24(3):209–228.

Barco, Carolina. 2013. Misión para el fortalecimiento del sistema de ciudades de Colombia. Documento técnico de soporte política del sistema de ciudades de Colombia. Report for the Colombian National Planning Ministry (Departamento Nacional de Planeación, Dirección de Desarrollo Urbano).

Baum-Snow, Nathaniel. 2007. Did highways cause suburbanization? *Quarterly Journal of Economics* 122(2):775–805.

Broda, Christian and David E. Weinstein. 2006. Globalization and the gains from variety. *Quarterly Journal of Economics* 121(2):541–585.

Bushnell, David. 1993. *The Making of Modern Colombia: A Nation in Spite of Itself*. Berkeley (CA): University of California Press.

Chandra, Amitabh and Eric Thompson. 2000. Does public infractsructure affect economic activity? Evidence from the rural interstate highway system. *Regional Science and Urban Economics* 30(4):457–490.

Disdier, Anne-Célia and Keith Head. 2008. The puzzling persistence of the distance effect on bilateral trade. *Review of Economics and Statistics* 90(1):37–48.

Donaldson, David. 2014. Railroads of the Raj: Estimating the impact of transportation infrastructure. *American Economic Review* (forthcoming).

Dube, Oeindrila and Juan F. Vargas. 2013. Commodity price shocks and civil conflict: Evidence from Colombia. *Review of Economic Studies* 80(4):1384–1421.

Duranton, Gilles. 2013. Agglomeration effects in Colombia. Processed, Wharton School, University of Pennsylvania.

Duranton, Gilles, Peter M. Morrow, and Matthew A. Turner. 2014. Roads and trade: Evidence from the US. *Review of Economic Studies* 81(2):681–724.

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

Fernald, John G. 1999. Roads to prosperity? Assessing the link between public capital and productivity. *American Economic Review* 89(3):619–638.

Gibbons, Stephen, Teemu Lyytikäinen, Henry Overman, and Rosa Sanchis-Guarner. 2012. New road infrastructure: The effects on firms. Processed, London School of Economics.

Michaels, Guy. 2008. The effect of trade on the demand for skill - Evidence from the Interstate Highway System. *Review of Economics and Statistics* 90(4):683–701.

Redding, Stephen and Anthony J. Venables. 2004. Economic geography and international inequality. *Journal of International Economics* 62(1):63–82.

Redding, Stephen J. and Matthew A. Turner. 2014. Transportation costs and the spatial organization of economic activity. *Handbook of Regional and Urban Economics*, forthcoming.

Roda, Pablo, Francisco Perdomo, and Jorge Sánchez. 2012. *Conectividad interurbana en Colombia*. Report for the Colombian National Planning Ministry (Departamento Nacional de Planeación, Dirección de Desarrollo Urbano).

Samad, Taimur, Nancy Lozano-Gracia, and Alexandra Panman. 2012. *Colombia Urbanization Review. Amplifying the Gains from the Urban Transition*. Washington (DC): World Bank.

Stock, James H. and Motohiro Yogo. 2005. Testing for weak instruments in linear IV regression. In Donald W.K. Andrews and James H. Stock (eds.) *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*. Cambridge: Cambridge University Press, 80–108.

Volpe-Martincus, Christian, Jerónimo Carballo, and Ana Cusolito. 2012. Routes, exports, and employment in developing countries: Following the trace of the Inca roads. Processed, Inter-American Development Bank.

Appendix A. Supplementary results

Table 16: First-step results, trade distances over 100 kilometres

	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) TSLS	(8) TSLS
Panel A. Dependent variable: (log) Weight of bilateral trade flows.								
log(driving time)	-0.47[a]			-1.23[a]	-0.92	-0.31	-0.47[a]	-25.2
	(0.051)			(0.25)	(2.44)	(0.27)	(0.049)	(42.3)
log(road distance)		-0.42[a]				-0.24		
		(0.045)				(0.33)		
log(Euclid. dist.)			-0.37[a]			0.097		
			(0.043)			(0.17)		
log(driving time)^2				0.21[a]	-0.38			19.8
				(0.068)	(2.20)			(35.4)
log(driving time)^3					0.33			-6.67
					(0.82)			(12.3)
log(driving time)^4					-0.056			0.81
					(0.11)			(1.52)
R^2	0.45	0.45	0.45	0.45	0.45	0.45	-	-
First stage stat.							20,732	2.94
Panel B. Dependent variable: (log) Value of bilateral trade flows.								
log(driving time)	-0.54[a]			-0.98[a]	1.59	-0.39	-0.53[a]	-33.2
	(0.058)			(0.29)	(2.98)	(0.31)	(0.055)	(26.0)
log(road distance)		-0.47[a]				-0.29		
		(0.051)				(0.37)		
log(Euclid. dist.)			-0.42[a]			0.15		
			(0.049)			(0.19)		
log(driving time)^2				0.12	-2.59			26.2
				(0.078)	(2.69)			(21.8)
log(driving time)^3					1.14			-8.78
					(1.01)			(7.61)
log(driving time)^4					-0.16			1.05
					(0.13)			(0.94)
R^2	0.41	0.41	0.41	0.41	0.42	0.41	-	-
First stage stat.							20,041	2.86

Notes: All regressions include importer and exporter fixed effects for all cities. Both panels report the same regressions but use different dependent variables: weight of trade flows for panel A and value of trade flows for panel B. The regressions are based on 6,150 observations in panel A and 6,021 observations in panel B. In columns 7 and 8, travel time variables are instrumented by their corresponding Euclidian distance terms. Robust standard errors in parentheses. ^{a}, ^{b}, ^{c}: significant at 1%, 5%, 10%.

Table 17: First-step results, trade in value with distance effects by product groups

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	OLS	OLS	OLS	OLS	OLS	TSLS	TSLS
Dependent variable: (log) value of bilateral trade flows.								
log(driving time)	-0.49[a]			-1.07[a]	-1.14[a]	-0.16	-0.49[a]	-1.52[a]
	(0.044)			(0.11)	(0.14)	(0.21)	(0.043)	(0.18)
log(road distance)		-0.43[a]				-0.48[c]		
		(0.038)				(0.29)		
log(Euclid. dist.)			-0.40[a]			0.20		
			(0.038)			(0.19)		
log(driving time)^2				0.22[a]	0.087			0.034
				(0.037)	(0.12)			(0.26)
log(driving time)^3					0.13			0.35[c]
					(0.088)			(0.19)
log(driving time)^4					-0.029			-0.083[b]
					(0.021)			(0.039)
Agriculture×dist.	0.13[b]	0.11[b]	0.093[b]	0.37[a]	0.42[b]	0.055	0.12[b]	0.75[a]
	(0.050)	(0.045)	(0.044)	(0.13)	(0.17)	(0.20)	(0.051)	(0.22)
Food×dist.	0.21[a]	0.17[a]	0.16[a]	0.40[a]	0.43[b]	0.21	0.21[a]	0.79[a]
	(0.056)	(0.049)	(0.048)	(0.14)	(0.19)	(0.22)	(0.056)	(0.24)
Mineral×dist.	0.19[c]	0.14	0.088	0.69[b]	0.65	0.27	0.12	0.59
	(0.11)	(0.091)	(0.088)	(0.32)	(0.43)	(0.43)	(0.10)	(0.50)
R^2	0.36	0.36	0.36	0.36	0.36	0.36	-	-
First stage stat.							4,425	2.68

Notes: All regressions include importer and exporter fixed effects for all cities. The regressions are based on 11,329 observations. In column 6, the interaction terms are only reported for the interactions with travel time. In columns 7 and 8, travel time variables are instrumented by their corresponding straight-line distance terms. Robust standard errors in parentheses. [a], [b], [c]: significant at 1%, 5%, 10%.

Table 18: Second-step results, LIML for importer effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Trade in weight and instruments from the 1938 road network.								
Road index	0.44[a]	0.38	0.18	0.058	0.14	-0.020	0.26[c]	-0.0096
	(0.14)	(0.34)	(0.14)	(0.17)	(0.14)	(0.15)	(0.14)	(0.14)
Controls								
2010 population	N	Y	Y	Y	Y	Y	Y	Y
Market access	N	Y	Y	Y	Y	Y	Y	Y
2011 wage	N	N	Y	Y	Y	Y	N	N
Past populations	N	N	N	N	Y	Y	Y	Y
Region indicators	N	N	N	Y	N	Y	N	Y
Overid. p-value	0.89	0.16	0.33	0.55	0.062	0.087	0.018	0.074
First-stage stat.	17.8	6.05	9.94	5.37	9.90	5.83	10.5	6.52
Observations	134	134	119	115	96	92	102	98
Panel B. Trade in value and instruments from the 1938 road network.								
Road index	0.49[a]	0.41	0.18	0.093	0.14	0.013	0.25[c]	0.032
	(0.15)	(0.34)	(0.14)	(0.17)	(0.14)	(0.16)	(0.13)	(0.15)
Overid. p-value	1.00	0.18	0.34	0.50	0.069	0.092	0.040	0.12
First-stage stat.	17.8	6.23	10.2	5.46	10.1	5.93	10.6	6.61
Observations	134	134	119	115	96	92	102	98

Notes: All regressions include a constant. Robust standard errors in parentheses unless otherwise mentioned. [a], [b], [c]: significant at 1%, 5%, 10%. Instruments are the number of kilometers of 1938 roads and distance to the nearest 1938 road.

Table 19: Second-step results, LIML for exporter effects when controlling for product groups

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Trade in weight and instruments from the 1938 road network.								
Road index	0.41[a]	0.35	0.24[b]	0.12	0.13	0.031	0.14	0.021
	(0.12)	(0.22)	(0.12)	(0.14)	(0.15)	(0.18)	(0.12)	(0.15)
Controls								
2010 population	N	Y	Y	Y	Y	Y	Y	Y
Market access	N	Y	Y	Y	Y	Y	Y	Y
2011 wage	N	N	Y	Y	Y	Y	N	N
Past populations	N	N	N	N	Y	Y	Y	Y
Region indicators	N	N	N	Y	N	Y	N	Y
Overid. p-value	0.65	0.19	0.11	0.14	0.16	0.15	0.31	0.39
First-stage stat.	17.8	6.11	10.7	6.49	13.1	7.46	14.8	9.03
Observations	134	134	119	115	96	92	102	98
Panel B. Trade in value and instruments from the 1938 road network.								
Road index	0.43[a]	0.38[c]	0.35[a]	0.27	0.27	0.19	0.28[b]	0.19
	(0.12)	(0.22)	(0.13)	(0.17)	(0.18)	(0.24)	(0.13)	(0.17)
Overid. p-value	0.67	0.18	0.063	0.048	0.026	0.020	0.079	0.061
First-stage stat.	17.8	6.10	10.7	6.47	13.1	7.42	14.7	8.95
Observations	134	134	119	115	96	92	102	98

Notes: All regressions include a constant. Robust standard errors in parentheses unless otherwise mentioned. [a], [b], [c]: significant at 1%, 5%, 10%. Instruments are the number of kilometers of 1938 roads and distance to the nearest 1938 road.