

# DISCUSSION PAPER SERIES

No. 10411

**INDUSTRY LOCATION AND WAGES: THE  
ROLE OF MARKET SIZE AND  
ACCESSIBILITY IN TRADING NETWORKS**

Javier Barbero, Kristian Behrens and Jose L. Zofio

***INTERNATIONAL TRADE AND  
REGIONAL ECONOMICS***



**Centre for Economic Policy Research**

# INDUSTRY LOCATION AND WAGES: THE ROLE OF MARKET SIZE AND ACCESSIBILITY IN TRADING NETWORKS

*Javier Barbero, Kristian Behrens and Jose L. Zofio*

Discussion Paper No. 10411

February 2015

Submitted 02 February 2015

Centre for Economic Policy Research  
77 Bastwick Street, London EC1V 3PZ, UK

Tel: (44 20) 7183 8801

[www.cepr.org](http://www.cepr.org)

This Discussion Paper is issued under the auspices of the Centre's research programme in **INTERNATIONAL TRADE AND REGIONAL ECONOMICS**. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as an educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Javier Barbero, Kristian Behrens and Jose L. Zofio

# INDUSTRY LOCATION AND WAGES: THE ROLE OF MARKET SIZE AND ACCESSIBILITY IN TRADING NETWORKS

## Abstract

We investigate the geographical distribution of economic activity and wages in a general equilibrium model with many asymmetric regions and costly trade. As shown by extensive simulations on random networks, local market size better explains a region's industry share, whereas accessibility better explains a region's wage. The correlation between equilibrium wages and industry shares is low, thus suggesting that the two variables operate largely independently. The model replicates well the spatial distribution of industry using Spanish data, yet overpredict changes in that distribution due to changes in 'generalized transport costs'. The latter had only small impacts on changes in the geographical distribution of economic activity in Spain from 1980 to 2007.

JEL Classification: C63, F12 and R12

Keywords: generalized transport costs, industry location, size, trading networks and wages

Javier Barbero [javier.barbero@uam.es](mailto:javier.barbero@uam.es)  
*Universidad Autónoma de Madrid*

Kristian Behrens [behrens.kristian@uqam.ca](mailto:behrens.kristian@uqam.ca)  
*Université du Québec à Montréal (ESG-UQAM), National Research University and CEPR*

Jose L. Zofio [jose.zofio@uam.es](mailto:jose.zofio@uam.es)  
*Universidad Autónoma de Madrid*

# 1 Introduction

Do market size and accessibility matter for industry location and wages? This question has attracted substantial attention in the literature ever since Krugman's (1980) and Helpman and Krugman's (1985) seminal contributions to new trade theory. The answer is 'yes', at least in simple models. It has indeed been shown that, in a world with increasing returns and costly trade, market size and 'accessibility' are locational advantages that significantly influence the geographical distribution of industry and regional factor prices.<sup>1</sup>

Despite those fundamental insights of new trade theory and new economic geography, it is fair to say that the models in which the results have been derived rely on a number of highly restrictive assumptions. Those assumptions include, among others: (i) the existence of a costlessly tradable good; (ii) a single production factor; (iii) constant elasticity of substitution (CES) preferences; (iv) two industries only, with one producing a homogeneous good; and (v) two locations only. Though necessary to derive clear-cut results, those assumptions imply that little is known about the robustness of the results and on how they can guide empirical analysis.

Conscious of those limitations, and of the fact that a better understanding is required to push further empirical work on that topic, much subsequent work has started to relax some of those assumptions. First, Ottaviano and Thisse (2004), Picard and Zeng (2005), Zeng and Kikuchi (2009), Baldwin, Martin, Forslid, Ottaviano, and Robert-Nicoud (2003), Head, Mayer, and Ries (2002), and Yu (2005), among others, have shown that the basic insights of 'home market effects' (HME) generalize to other preference structures – such as quadratic-linear preferences – or other market structures – such as oligopolistic competition. Second, Davis (1998) and Picard and Zeng (2005) have shown that the effect of market size on industry location is strongly dampened (or even disappears) when the homogeneous good is not costlessly tradable. Davis (1998), in particular, shows that when trading the homogeneous good is as costly as trading the differentiated good, market size has no longer any bearing on regional specialization. This is also one basic message of Hanson and Xiang (2004), who argue that – in the absence of a costlessly tradable good – not all increasing returns sectors can be disproportionately present in one region, i.e., display 'home market effects'. Zeng and Kikuchi (2009), and Takahashi, Takatsuka, and Zeng (2013) derive analytical results in the case without factor price equalization (FPE), but only with two regions.<sup>2</sup> Behrens, Lamorgese, Ottaviano, and

---

<sup>1</sup>The so-called 'home market effect' (HME) quickly became a key building block of New Trade Theory (NTT) first, and New Economic Geography (NEG) later (Krugman, 1991; Ottaviano, Tabuchi, and Thisse, 2002). It also attracted much attention in the empirical trade literature (e.g., Davis and Weinstein, 2003; Hanson and Xiang, 2004), because it potentially allowed to investigate the role of market size in shaping industry structure and trade.

<sup>2</sup>See also Takatsuka and Zeng (2012) for another model with trade costs and two countries. As in Picard and Zeng (2005), all those models build on quadratic quasi-linear preferences following Ottaviano, Tabuchi, and Thisse (2002). These authors show that trade costs in the homogeneous good do not 'obscure' the HME, but the results are difficult to compare with those of the literature since there are no income effects in those models.

Tabuchi (2009) use a ‘hybrid’ approach, where trading the homogenous good is costless, but where exogenous Ricardian differences in labor productivity in the homogeneous sector across countries create exogenous wage differences. Though conceptually simple and applicable to multiple countries, that approach does not allow for endogenous wage changes in response to changes in economic fundamentals. Last, turning to multi-country extensions of those models, Behrens, Lamorgese, Ottaviano, and Tabuchi (2007, 2009) derive results for models with more than two countries. They show that the topology of the trading network matters for several of the results, and that the impact of market size on industry location arises only when differences in factor costs and in accessibility to markets are adequately controlled for. While being empirically very important, multi-location extensions of new trade models to arbitrary geographical structures have been very rare in the literature until now.<sup>3</sup>

While all of the foregoing contributions shed some light on the role of market size and accessibility on industry location and wages, what is missing to date is more systematic evidence for what happens in more ‘realistic settings’ where several of the basic assumptions are relaxed simultaneously. To the best of our knowledge, there has been no systematic investigation when there are multiple locations, several industries, and costly trade for all goods. This paper addresses precisely these issues. As there is no hope to obtain clear-cut analytical results in the general case, we instead resort to systematic numerical simulations. More precisely, we simulate the equilibria of two different models using a large number of randomly generated networks with a large number of regions. We then run simple regressions to extract the essence of the ‘comparative static’ results that are out of reach of pencil-and-paper analysis. In a nutshell, our research strategy is to combine theory and numerical analysis to: (i) first prove some theorems in ‘toy models’; (ii) then solve large-scale models by numerical analysis; (iii) then run a detailed statistical analysis of the numerical results, very much like engineers or physicists do; and (iv) finally confront the models with real data to use if for simulation purposes.

Our key findings can be summarized as follows. First, absolute local market size – as measured by population – and accessibility – as measured by centrality in the trading network – are crucial in explaining a region’s wage. This result is due to the fact that absolute size and accessibility affect all industries in similar ways, i.e., constitute a region’s *absolute advantage*. The effect is stronger and more systematic in models where all sectors are subject to transport costs and exhibit increasing returns to scale. Second the relative local market size of industries (as measured by their consumer expenditure shares) is crucial in explaining a region’s industrial composition. This result is due to the fact that relative spending patterns

---

<sup>3</sup>There are many other dimensions along which the basic models have been extended. Without being exhaustive, we can mention Zeng and Kikuchi (2009), who provide an alternative model based on a footloose-capital specification with two production factors: labor, for the variable cost, and capital for the fixed cost. Behrens and Picard (2007, 2011) show that ‘home market effects’ get weaker or can even be reversed in the presence of multi-plant firms, or when trade imbalances translate into higher freight rates.

do not affect all industries in similar ways, i.e., constitute a region's *comparative advantage*. In a nutshell – and in line with Ricardian trade theory – absolute advantage translates into wages, whereas comparative advantage maps into specialization patterns. Third, the correlation between equilibrium wages and equilibrium industry shares is rather low, thus suggesting that both variables operate largely independently.

We then apply the models to Spanish data. Using ‘Generalized Transport Costs’ between regions as a measure of trade frictions, we find that the models generally predict well the distribution of industries, yet predict less well wages. A formal test does not allow to reject the null hypothesis that the industry distribution predicted by the models is the same than that observed in the data. We then use the calibrated model for Spain to run two counterfactual exercises, the aim of which is to disentangle the impact of changes in accessibility and changes in market size on regional industry shares and wages. Holding population fixed at 1980 levels, we find that changes in transport costs between 1980 and 2007 do not explain much of the increase in regional inequalities observed in Spain during that period. The change in inequality is much better captured when we hold transport costs fixed at 1980 levels and consider changes in population shares between 1980 and 2007. Although the simulated models capture the qualitative trend towards more regional inequality in Spain, they also tend to significantly overpredict the increase in polarization observed between 1980 and 2007.

The remainder of the paper is organized as follows. Section 2 develops two different new trade ‘toy models’: one with a single differentiated industry and a homogeneous good industry; and one with two differentiated industries. In both models, trade is costly and factor prices are endogenous. In Section 3, we extend the models to a larger scale and discuss a set of numerical results obtained from simulating those two models for a large number of random networks. We then present, in Section 4, an application to the case of Spanish regional data, as well as results from two counterfactuals. Finally, Section 5 concludes. Technical details are relegated to an extensive set of appendices.

## 2 Models

We develop two models within which we analyze the geographical distribution of economic activity and wages. In both models, there are  $M \geq 2$  regions subscripted by  $i = 1, 2, \dots, M$ . Each region is endowed with  $L_i$  immobile workers-consumers. The total population in the economy is fixed at  $L \equiv \sum_i L_i$ . Labor is the only production factor, i.e., we abstract from comparative advantage across regions.

## 2.1 Model 1: One differentiated sector and one homogeneous sector

Our first model builds on Helpman and Krugman (1985) and its multi-location extensions by Behrens *et al.* (2007, 2009). There is one increasing returns to scale (IRS) sector with monopolistic competition that produces a continuum of varieties of a horizontally differentiated good; and one constant returns to scale (CRS) sector with perfect competition that produces a homogeneous good. In the differentiated sector, the combination of IRS, costless product differentiation, and the absence of scope economies yields a one-to-one equilibrium relationship between firms and varieties.

### 2.1.1 Preferences and demands

Preferences of a representative consumer in region  $j$  are given by:

$$U_j = H_j^{1-\mu} D_j^\mu, \quad (1)$$

where  $H_j$  stands for the consumption of the homogeneous good; where  $D_j$  is an aggregate of the varieties of the differentiated good; and where  $0 < \mu < 1$  is the income share spent on the differentiated good. We assume that  $D_j$  is given by a CES subutility function

$$D_j = \left[ \sum_i \int_{\Omega_i} d_{ij}(\omega)^{(\sigma-1)/\sigma} d\omega \right]^{\frac{\sigma}{\sigma-1}},$$

where  $d_{ij}(\omega)$  is the individual consumption in region  $j$  of variety  $\omega$  produced in region  $i$ ; and where  $\Omega_i$  is the set of varieties produced in  $i$ . The parameter  $\sigma > 1$  measures the elasticity of substitution between any two varieties. Let  $p_j^H$  denote the price of the homogeneous good in region  $j$  and  $p_{ij}(\omega)$  the price of variety  $\omega$  produced in region  $i$  and consumed in region  $j$ . Let  $w_j$  denote the wage in region  $j$ . Maximizing (1) subject to the budget constraint  $p_j^H H_j + \sum_i \int_{\Omega_i} p_{ij}(\omega) d_{ij}(\omega) d\omega = w_j$  yields the following individual demands:

$$d_{ij}(\omega) = \frac{p_{ij}(\omega)^{-\sigma}}{\mathbb{P}_j^{1-\sigma}} \mu w_j \quad \text{and} \quad H_j = \frac{(1-\mu)w_j}{p_j^H}, \quad (2)$$

where  $\mathbb{P}_j$  is the CES price index in region  $j$ , given by

$$\mathbb{P}_j = \left[ \sum_i \int_{\Omega_i} p_{ij}(\omega)^{1-\sigma} d\omega \right]^{\frac{1}{1-\sigma}}. \quad (3)$$

### 2.1.2 Differentiated good

We first explain the workings of the IRS industry. Technology is assumed to be identical across firms and regions, therefore implying that firms differ only by the variety they produce and the

region they are located in. Since varieties enter preferences in a symmetric way, we henceforth suppress the variety index  $\omega$  to alleviate notation. Production of any variety involves a fixed labor requirement,  $F$ , and a constant marginal labor requirement,  $c$ . Denote by  $x_{ij}$  the amount of a variety produced in  $i$  and shipped to  $j$ . The total labor requirement for producing output  $x_i \equiv \sum_j x_{ij}$  is given by  $l_i = F + cx_i$ .

Trade in the differentiated good is costly. Following standard practice we assume that trade cost are of the *iceberg* form:  $\tau_{ij} \geq 1$  units must be dispatched from region  $i$  in order for one unit to arrive in region  $j$ . We further assume that trade costs are symmetric, i.e.,  $\tau_{ij} = \tau_{ji}$ .<sup>4</sup> Using the demands (2), each firm in  $i$  maximizes its profit

$$\pi_i = \sum_j (p_{ij} - cw_i\tau_{ij}) L_j \frac{p_{ij}^{-\sigma}}{\mathbb{P}_j^{1-\sigma}} \mu w_j - Fw_i \quad (4)$$

with respect to all its prices  $p_{ij}$ , taking the price indices  $\mathbb{P}_j$  and the wages  $w_j$  as given. Because of CES preferences, profit-maximizing prices have constant markups

$$p_{ij} = \frac{\sigma}{\sigma - 1} cw_i\tau_{ij}. \quad (5)$$

We denote by  $n_i$  the endogenously determined mass of firms located in  $i$ , and by  $N \equiv \sum_i n_i$  the total mass of firms in the economy. We also denote by  $\lambda_i \equiv n_i/N$  the share of firms in region  $i$ .

Because of iceberg trade costs, a firm in region  $i$  has to produce  $x_{ij} \equiv L_j d_{ij} \tau_{ij}$  units to satisfy aggregate demand in region  $j$ . Free entry and exit imply that profits are non-positive in equilibrium which, using (4) and the pricing rule (5), yields the standard condition

$$x_i \equiv \sum_j L_j d_{ij} \tau_{ij} \leq \frac{F(\sigma - 1)}{c}. \quad (6)$$

Let  $\phi_{ij} \equiv \tau_{ij}^{1-\sigma} \in [0, 1]$  denote the ‘freeness of trade’ in the differentiated good between regions  $i$  and  $j$ . Inserting the demand (2) and the price index (3) into (6), multiplying both sides by  $p_{ij}$ , and using the prices (5), we get the wage equations

$$\sum_j \frac{w_i^{-\sigma} w_j \phi_{ij} L_j}{\sum_k w_k^{1-\sigma} \phi_{kj} n_k} \leq \frac{\sigma F}{\mu}. \quad (7)$$

Dividing both sides by the total population,  $L$ , letting  $\theta_j \equiv L_j/L$ , and choosing – without loss of generality – units for  $F$  such that  $F \equiv \mu L/\sigma$ , we can rewrite (7) as follows:

$$\text{RMP}_i \equiv \sum_j \frac{w_i^{-\sigma} w_j \phi_{ij} \theta_j}{\sum_k w_k^{1-\sigma} \phi_{kj} n_k} \leq 1, \quad (8)$$

---

<sup>4</sup>This assumption is not crucial but makes our life easier in terms of modeling. We relax it later in Section 4 when applying our model to Spanish regions.



where  $RMP_i$  stands for the *real market potential* of region  $i$  (Head and Mayer, 2004). The number of workers employed in the differentiated industry of region  $i$ , when it has  $n_i$  firms, is

$$L_i^D \equiv n_i l_i = n_i (F + cx_i) = n_i \mu L, \quad (9)$$

where we have made use of our normalization of  $F$ .

### 2.1.3 Homogeneous good

We next explain the workings of the perfectly competitive CRS industry. We assume that technology is the same in all regions. Without loss of generality, we normalize the unit labor requirement to one. Perfect competition implies marginal cost pricing. Given  $L_i^D$  workers employed in the differentiated good industry, the number of workers employed in the homogeneous sector equals  $L_i^H \equiv L_i - L_i^D$ . Inserting (9) into that expression, we can rewrite the number of workers in the homogeneous sector as follows:

$$L_i^H = L_i - n_i \mu L. \quad (10)$$

Note that (10) need not be positive, i.e., some regions may specialize in the production of the differentiated good only.

We assume that trading the homogeneous good is costly.<sup>5</sup> Hence, factor price equalization (FPE) does not hold in general and the world mass of firms in the differentiated industry is no longer constant.<sup>6</sup> The price of the homogeneous good produced in  $i$  and delivered to  $j$  equals its marginal cost of production, the wage  $w_i$ , times the trade cost  $\tau_{ij}^H$  between regions  $i$  and  $j$ :  $p_{ij}^H = w_i \tau_{ij}^H \equiv w_i \xi \tau_{ij}$ , where  $\xi > 0$  is a parameter that captures the *relative cost of trading the homogeneous good compared to the differentiated good*. If  $\xi = 1$ , there are no cost differences. When  $\xi > 1$ , trading the homogeneous good is more costly than trading the differentiated good, and vice versa when  $\xi < 1$ . In what follows, we set  $\xi < 1$  because in the opposite case there is no trade in the homogeneous good so that the only equilibrium is one where industry shares are proportional to the size of the local market (Davis, 1998).<sup>7</sup>

Because good  $H$  is homogeneous and can be produced in, and imported from, any region, its price in region  $i$  must be the lowest one that can be secured from any source:

$$p_i^H = \min_k \{w_k \xi \tau_{ki}\}. \quad (11)$$

---

<sup>5</sup>See Appendix A for a discussion of the case with costless trade of the homogeneous good. In that Appendix, we also explain why we disregard this case in what follows.

<sup>6</sup>The total mass of firms,  $N$ , varies with the spatial structure of the economy when there is costly trade in the homogeneous good (see, e.g., Takatsuka and Zeng, 2012). Hence, (8) cannot be generally expressed in the usual share notation  $\lambda_i$  with respect to firms, which explains the presence of  $n_k$  in that expression.

<sup>7</sup>There is of course still two-way trade in the differentiated good and the wages adjust to balance that trade. However, our focus is on industry structure and wages. The former cannot be meaningfully analyzed when we assume that  $\xi \geq 1$ , whereas the latter cannot be meaningfully analyzed if we assume that there is free trade in the homogeneous good.

Demand for the homogeneous good is given by (2), while supply is determined by the domestic production for the local market,  $X_{ii}$ , and the sum of imports. Let  $X_{ji}$  denote the imports of the homogeneous good from region  $j$ . Market clearing for the homogeneous good in region  $i$  requires that:

$$\frac{(1 - \mu) w_i L_i}{p_i^H} = X_{ii} + \sum_{j \neq i} X_{ji}. \quad (12)$$

Dividing the foregoing expression by the total population,  $L$ , and using the price (11), we can write (12) in terms of population shares, production, and imports:

$$\frac{(1 - \mu) w_i \theta_i}{\min_k \{w_k \xi \tau_{ki}\}} = \tilde{X}_{ii} + \sum_{j \neq i} \tilde{X}_{ji}, \quad (13)$$

where  $\tilde{X}_{ii} \equiv X_{ii}/L$ , and  $\tilde{X}_{ji} \equiv X_{ji}/L$  denote per capita variables. Labor market clearing in region  $i$  then requires that  $L_i^H = L_i - n_i \mu L = \xi (\tau_{ii} X_{ii} + \sum_{j \neq i} \tau_{ij} X_{ij})$ . Since  $L_i = \theta_i L$ , we can rewrite the foregoing condition in per capita terms as follows:

$$\theta_i - n_i \mu = \xi \left( \tau_{ii} \tilde{X}_{ii} + \sum_{j \neq i} \tau_{ij} \tilde{X}_{ij} \right). \quad (14)$$

Because of perfect competition, the homogeneous good will not be simultaneously imported and exported by the same region. Hence, it must be that

$$\tilde{X}_{ij} = \begin{cases} > 0 & \text{if } w_i \tau_{ij} \leq \min_k \{w_k \tau_{kj}\} \\ = 0 & \text{otherwise.} \end{cases}$$

This latter condition can be rewritten equivalently in complementary slackness terms as

$$\tilde{X}_{ij} \cdot \left[ w_i \tau_{ij} - \min_k \{w_k \tau_{kj}\} \right] = 0 \quad \text{and} \quad \tilde{X}_{ij} \geq 0, \quad \forall j = 1, 2, \dots, M. \quad (15)$$

#### 2.1.4 Equilibrium

An equilibrium in the IRS sector is such that the real market potential (8) is equal to one in all regions with a positive measure of firms, and less than one for regions devoid of firms. If all regions have a positive measure of firms, we obtain an *interior equilibrium*, whereas if there are some regions without firms we get a *corner equilibrium*. Formally, an equilibrium is defined as:

$$\begin{aligned} \text{RMP}_i &= 1 & \text{if } n_i^* > 0 \\ \text{RMP}_i &\leq 1 & \text{if } n_i^* = 0. \end{aligned} \quad (16)$$

Using notation in terms of complementary slackness, this implies that  $n_i^* \cdot (\text{RMP}_i - 1) = 0$  and  $n_i^* \geq 0$  for all regions. In addition to the zero profit free entry condition (16), the market clearing conditions (14) for the homogeneous good must hold for all regions at the equilibrium wages  $w_i$ . Conditions (13), (14), (15), and (16) define a system of  $3M + M(M - 1)$  equations in as many unknowns – the firm masses  $n_i$ , the wages  $w_i$ , the per capita domestic supplies  $\tilde{X}_{ii}$ , and the per capita imports  $\tilde{X}_{ij}$ .

## 2.2 Model 2: Two differentiated sectors

Our second model builds on Krugman (1980) and Behrens and Ottaviano (2011). There are two IRS sectors with CES monopolistic competition.<sup>8</sup> Regional market sizes differ both because of the numbers of consumers and because consumers have different spending patterns for the two goods. In such a setting, we can look at how differences in *absolute market sizes* – the population shares  $\theta_i$  – and differences in *relative market sizes* – the expenditure shares  $\mu_i$  – affect wages and the location patterns of industries.

### 2.2.1 Preferences and demands

The basic setup is the same as in Section 2.1, except that there are now two CES sectors and no homogeneous sector. Preferences of a representative consumer in region  $j$  are given by:

$$U_j = D_{1j}^{\mu_{1j}} D_{2j}^{\mu_{2j}}, \quad (17)$$

where  $D_{sj}$  stands for the CES consumption aggregate in sector  $s$  in region  $j$ ; and  $0 < \mu_{sj} < 1$  is the *region-specific* income shares for sector  $s$ . With two sectors,  $\mu_{sj}$  is equal to  $\mu_j$  in sector 1 and to  $1 - \mu_j$  in sector 2. Since expenditure shares are region specific, the relative consumption patterns differ across regions. Hence, market sizes differ due to spending patterns on top of differences in regional population sizes.

The aggregator for consumption of the differentiated good,  $D_{sj}$ , is as follows:

$$D_{sj} = \left[ \sum_i \int_{\Omega_{si}} d_{sij}(\omega)^{(\sigma-1)/\sigma} d\omega \right]^{\frac{\sigma}{\sigma-1}},$$

where  $d_{sij}(\omega)$  is the individual consumption in region  $j$  of sector- $s$  variety  $\omega$  produced in region  $i$ ; and where  $\Omega_{si}$  is the set of sector- $s$  varieties produced in  $i$ . For simplicity, we assume that the elasticity of substitution between any two varieties,  $\sigma$ , is the same in both sectors.<sup>9</sup> Let  $p_{sij}(\omega)$  denote the price of sector- $s$  variety  $\omega$  produced in  $i$  and consumed in  $j$ ; and let  $w_j$  denote the wage in region  $j$ . Maximizing (17) subject to the budget constraint  $\sum_i \left[ \int_{\Omega_{1i}} p_{1ij}(\omega) d_{1ij}(\omega) d\omega + \int_{\Omega_{2i}} p_{2ij}(\omega) d_{2ij}(\omega) d\omega \right] = w_j$  yields the following individual demands:

$$d_{sij}(\omega) = \frac{p_{sij}(\omega)^{-\sigma}}{\mathbb{P}_{sj}^{1-\sigma}} \mu_{sj} w_j, \quad \text{where} \quad \mathbb{P}_{sj} = \left[ \sum_i \int_{\Omega_{si}} p_{sij}(\omega)^{1-\sigma} d\omega \right]^{\frac{1}{1-\sigma}} \quad (18)$$

<sup>8</sup>Hanson and Xiang (2004) develop a model with a continuum of sectors, but their focus is on two regions only. In this section, we take a complementary approach: we focus on two sectors only, but consider a large number of regions to look at industry location and wages.

<sup>9</sup>We could relax that assumption, but there is not much to be learned from that exercise. The same holds true for relaxing the assumption of identical technologies in the two sectors. Nevertheless, as explained in footnote 17 below, we have also studied the effects of alternative values of  $\sigma$  and expenditure patterns  $\mu_{sj}$  on industry shares,  $\lambda_{si}$ , and wages,  $w_i$ .

is the CES price index in sector  $s$  and region  $j$ .

### 2.2.2 Technology and trade

For simplicity, we assume that technology is the same in both sectors. As in Section 2.1, the total labor requirement for producing the output  $x_{si} \equiv \sum_j x_{sij}$  is given by  $l_{si} = F + cx_{si}$ . Trade in both differentiated goods is costly and trade cost are symmetric and of the iceberg form:  $\tau_{sij} = \tau_{sji} \geq 1$  units must be dispatched from region  $i$  in order for one unit of a sector- $s$  variety to arrive in region  $j$ . Using (18), a sector- $s$  firm in  $i$  maximizes profit

$$\pi_{si} = \sum_j (p_{sij} - cw_i \tau_{sij}) L_j \frac{p_{sij}^{-\sigma}}{\mathbb{P}_{sj}^{1-\sigma}} \mu_{sj} w_j - Fw_i \quad (19)$$

with respect to all its prices  $p_{sij}$ , taking the price indices  $\mathbb{P}_{sj}$  and the wages  $w_j$  as given. As before, profit-maximizing prices have constant markups:

$$p_{sij} = \frac{\sigma}{\sigma - 1} cw_i \tau_{sij}. \quad (20)$$

We denote by  $n_{si}$  the endogenously determined mass of sector- $s$  firms located in  $i$ , and by  $N_s \equiv \sum_i n_{si}$  the total mass of sector- $s$  firms in the economy. Last,  $\lambda_{si} \equiv n_{si}/N_s$  denotes the share of sector- $s$  firms in region  $i$ .

A firm in region  $i$  and sector  $s$  has to produce  $x_{sij} \equiv L_j d_{sij} \tau_{sij}$  units to satisfy aggregate demand in region  $j$ . Free entry and exit imply that profits are non-positive in equilibrium which, using the prices (20), yields again the standard free entry zero profit condition (6). Inserting the demands and the price index (18) into that expression, using the prices (20), and letting  $\phi_{sij} \equiv \tau_{sij}^{1-\sigma} \in [0, 1]$  denote the ‘freeness of trade’ in sector  $s$ , we get the wage equations:

$$\sum_j \frac{w_i^{-\sigma} w_j \phi_{sij} L_j \mu_{sj}}{\sum_k w_k^{1-\sigma} \phi_{skj} n_{sk}} \leq \sigma F. \quad (21)$$

Dividing both sides by world population,  $L$ , letting  $\theta_j \equiv L_j/L$  as before, and choosing without loss of generality units of  $F$  such that  $F = L/\sigma$ , we obtain the real market potential for sector- $s$  firms in region  $i$  as follows:

$$\text{RMP}_{si} \equiv \sum_j \frac{w_i^{-\sigma} w_j \phi_{sij} \theta_j \mu_{sj}}{\sum_k w_k^{1-\sigma} \phi_{skj} n_{sk}} \leq 1. \quad (22)$$

### 2.2.3 Equilibrium

Expressions (22) define  $2M$  conditions in the  $3M$  unknowns  $\{n_{1i}, n_{2i}, w_i\}$ , for  $i = 1, 2, \dots, M$ . To pin down the wages, we can impose either the labor market clearing conditions or the trade balance conditions. In what follows, we use the former as they are easier to handle given our

choices of normalization. Labor market clearing in  $i$  requires that  $L_i = n_{1i}(F + cx_{1i}) + n_{2i}(F + cx_{2i}) = L(n_{1i} + n_{2i})$ , where we have used the normalization of  $F$ . Hence,

$$\theta_i = n_{1i} + n_{2i}. \quad (23)$$

Conditions (22) and (23) can be solved for the equilibrium wages and industry shares. The total masses of firms in the two sectors in the economy,  $N_1 = \sum_i n_{1i}$  and  $N_2 = \sum_i n_{2i}$  are not constant and vary with the spatial distribution of demand and with the structure of the trading network. Note, of course, that the total mass of firms in both sectors in the world economy is equal to one:  $\sum_i (n_{1i} + n_{2i}) = \sum_i \theta_i = 1$  from (23).

To solve the model, we set  $w_1 \equiv 1$  by choice of numeraire. Focusing on two regions with symmetric trade costs and free intra-regional trade ( $\phi_{sii} = 1$  and  $\phi_{sij} = \phi$  for all  $i \neq j$ ), Behrens and Ottaviano (2011) have proven the following analytical results for two special cases: *absolute advantage*, i.e., when the spending patterns of the two regions are the same but when the regions differ by population size ( $\mu_{11} = \mu_{12}$  and  $\mu_{21} = \mu_{22}$ , but  $\theta_1 > \theta_2$ ); and *comparative advantage*, i.e., when spending patterns are anti-symmetric but when the regions have the same population size ( $\mu_{11} = \mu_{22}$  and  $\mu_{21} = \mu_{12}$ , but  $\theta_1 = \theta_2$ ). In those two polar cases, it can be shown that (see Behrens and Ottaviano, 2011, for the proofs):

**Proposition 1 (Pure ‘Comparative Advantage’)** *Assume that preferences are anti-symmetric across regions ( $\mu_{11} = \mu_{22}$  and  $\mu_{21} = \mu_{12}$ ), and that both regions are of the same size ( $\theta_1 = \theta_2$ ). The equilibrium is such that*

$$n_{11}^* = n_{22}^* = \frac{\mu(1 + \phi) - \phi}{2(1 - \phi)} \quad \text{and} \quad n_{21}^* = n_{12}^* = \frac{1 - \mu(1 + \phi)}{2(1 - \phi)}$$

*The equilibrium relative wage satisfies  $w_2^* = 1$ .*

**Proposition 2 (Pure ‘Absolute Advantage’)** *Assume that preferences are symmetric across regions ( $\mu_{11} = \mu_{12}$  and  $\mu_{21} = \mu_{22}$ ), and that region 1 has the larger market ( $\theta_1 > \theta_2$ ). The equilibrium is such that*

$$n_{1i}^* = \mu\theta_i \quad \text{and} \quad n_{2i}^* = (1 - \mu)\theta_i \quad (24)$$

*for  $i = 1, 2$ . The equilibrium relative wage satisfies  $0 < w_2^* < 1$ .*

In Proposition 1, each region is the larger market for one of the two goods. Hence, each region specializes in the production of the good for which it has a relatively larger local demand. In other words, relative differences in market sizes lead to different specialization patterns but do not affect factor prices. In the case of Proposition 2, one region is the larger market for both goods. In that case, the wage in the larger region must be higher because it offers a locational advantage for *both* industries. Clearly, this is akin to absolute advantage in a Ricardian sense and it is, therefore, capitalized into factor prices.

Of course, the two cases in Propositions 1 and 2 are extreme ones, and intermediate cases where both absolute and comparative advantage play a role should be considered. Furthermore, it is of interest to relax the assumption of just two regions and of symmetric trade costs to investigate also the interactions with ‘geography’. This is what we do using numerical simulations in the next section and Spanish data in Section 4.

### 3 Size and accessibility in random tree networks

It is virtually impossible to derive general analytical results in an arbitrary multi-region setting without FPE, because the equilibrium allocations of firms and wages are determined by a complex trade-off between a region’s market size and its accessibility in the trading network.<sup>10</sup> To nevertheless gain insights into how size and accessibility – as well as the whole structure of the trading network – influence the equilibrium, we resort to systematic numerical simulation. To this end, we proceed as follows.

First, we generate a random tree network with a random number of nodes (see Appendix B for details). The nodes are the regions, and the links between nodes represent the connections for shipping goods. Networks are generated incrementally either by having equal attachment probabilities for new nodes, or by using the Barabási and Albert (1999; henceforth BA) preferential attachment algorithm that generates networks which exhibit a ‘hub-and-spoke’ structure. Second, we assign a random population share,  $\theta_i$ , to each node  $i$  of the network.<sup>11</sup> In the case with two differentiated industries, we also randomly assign a region-specific expenditure share for each industry. Third, we solve the two models for their equilibria. We repeat this three-step process for a large number of randomly generated networks and then relate selected characteristics of the equilibria thus obtained to underlying networks characteristics. Doing so will allow us to gain more systematic insights into how size and accessibility interact to determine the regional allocation of firms and wages, and how those allocations depend on the economic model we have chosen.

We describe the numerical implementation in detail in Appendix C. In the following sec-

---

<sup>10</sup>One can derive analytical results by having multiple regions and a symmetric trading network, but this is in the end isomorphic to using just two regions and thus of no particular interest.

<sup>11</sup>Choosing ‘totally random’ networks – though providing an interesting benchmark case – is not fully satisfying because transportation networks are endogenous and obey certain rules. This is why we also derive results using networks that display a ‘hub-and-spoke’ structure to capture the empirical fact that some places are very well connected while others are very poorly connected (see, e.g., Xie and Levinsohn, 2008, for the case of the road network in Indiana). Observe that we assign  $\theta_i$  randomly, i.e., there is no systematic correlation between size and accessibility. The reason for that choice is that we want to study the distribution of industry as a function of size and accessibility separately. Introducing a systematic correlation between the two (though empirically relevant since larger places are better connected and since places that are better connected tend to grow larger; see Duranton and Turner, 2012) is not required for our analysis.

tions, we explore the results obtained for the two models.

### 3.1 Model 1: One differentiated sector and one homogeneous sector

We first compute simple correlations between the equilibrium masses of firms in the different regions ( $n_i^*$ ), their population shares ( $\theta_i$ ), and their centrality ( $C_i$ ). The latter is measured either by the *closeness centrality* (henceforth ‘closeness’, for short) or by the node’s degree. Following standard practice in the network literature (see, e.g., Freeman, 1979), closeness is defined as

$$C_i = \left[ \frac{\sum_j d_{ij}}{\min_k \{\sum_j d_{kj}\}} \right]^{-1}, \quad (25)$$

where  $d_{ij}$  denotes the length of the link – the distance – between nodes  $i$  and  $j$ . By definition, closeness varies between 0 and 1. ‘Degree’ is simply measured by the number of links of the node. Centrally located nodes have both a high value for closeness and for degree. This can be seen from the correlations in the top panel of Table 1. That panel shows that, as expected, size ( $\theta_i$ ) and accessibility (closeness $_i$  and degree $_i$ ) are positively linked to a region’s equilibrium industry share ( $\lambda_i^*$  or, alternatively,  $n_i^*$ ) and to a region’s wage ( $w_i^*$ ). The correlation is particularly strong for the degree measure of centrality. Observe also that size is more strongly linked to industry location, whereas accessibility is more strongly linked to wages. Put differently, size differences map into differences in industry structures, whereas accessibility differences translate into factor price difference. In general, however, the correlations with factor prices are weaker than the correlations with industry location.

Table 1: Simple correlations (Model 1).

	$\lambda_i^*$	$n_i^*$	$w_i^*$	$\theta_i$	closeness $_i$	degree $_i$
$\lambda_i^*$	1					
$n_i^*$	0.9987	1				
$w_i^*$	0.0849	0.0806	1			
$\theta_i$	0.8119	0.8065	0.0899	1		
closeness $_i$	0.2680	0.2693	0.1316	0.0134	1	
degree $_i$	0.3972	0.4023	0.1799	0.0135	0.7075	1
	CV( $\Lambda^*$ )	CV( $\mathbf{n}^*$ )	CV( $\mathbf{w}^*$ )	CV( $\theta$ )	CV(closeness)	CV(degree)
CV( $\Lambda^*$ )	1					
CV( $\mathbf{n}^*$ )	1	1				
CV( $\mathbf{w}^*$ )	-0.0082	-0.0082	1			
CV( $\theta$ )	0.3005	0.3005	0.1581	1		
CV(closeness)	0.0113	0.0113	0.2475	0.2156	1	
CV(degree)	0.3255	0.3255	-0.0728	-0.2254	-0.2215	1

Notes: Simple correlations for 100 random tree networks with a random number of 20 to 30 nodes. The top panel of the table gives correlations at the level of individual nodes (pooled across all 100 networks), whereas the bottom panel of the table gives correlations at the level of the whole network. The shares  $\lambda_i^*$  are given by  $\lambda_i^* = n_i^*/(\sum_j n_j^*)$ . CV denotes the coefficient of variation, whereas  $\Lambda^*$ ,  $\mathbf{n}^*$ ,  $\mathbf{w}^*$ , and  $\theta$  denote the equilibrium vectors of industry shares, masses of firms, wages, and market size, respectively.

The bottom panel of Table 1 displays the same correlations as in the top panel, but now between aggregate network statistics and the vectors of equilibrium outcomes. More precisely, it displays the correlations between the coefficients of variation (cv) of industry shares and wages (computed for each network at the node level), and the coefficients of variation of size and accessibility. As can be seen, dispersion in market sizes – as captured by a larger cv – is positively associated with dispersion of industry shares and wages. The same holds true for dispersion in the accessibility measures, with again a much stronger effect of degree as compared to closeness. It is finally of interest to note that the correlations between the equilibrium industry shares,  $\lambda_i^*$  (or the equilibrium masses of firms,  $n_i^*$ ) and the equilibrium wages – though positive – are fairly small (0.080 and 0.085, respectively). This result suggests that the two variables operate largely independently to determine the equilibrium.

To go beyond simple univariate correlations, we now run several ordinary least squares (OLS) regressions to estimate the partial effect of increasing market size or centrality of nodes on the equilibrium shares of manufacturing activity and the equilibrium wages, controlling for accessibility and for size. In Model 1, there are two endogenous variables that can be analyzed in the regressions: the equilibrium allocation of firms,  $\lambda_i^*$ , and the equilibrium wages,  $w_i^*$ .<sup>12</sup> Mirroring the two panels of Table 1, we start with an analysis at the level of the individual nodes, and turn then to an analysis at the level of the whole network.

### 3.1.1 Results for individual nodes

We regress the equilibrium shares of firms,  $\lambda_i^*$ , or the equilibrium wages,  $w_i^*$ , on measures of: (i) the node’s centrality, as given by either closeness or degree; and (ii) the node’s market size.<sup>13</sup> We perform a pooled analysis with both types of networks (based on preferential attachment, BA, or equal probabilities) – in which case we include a network dummy indicating the network type – and separate regressions for each type of network. Formally, we estimate

$$\lambda_i^* = \beta_0 + \beta_1 \text{centrality}_i + \beta_2 \theta_i + \text{network\_dummy}_i + \epsilon_i \quad (26)$$

$$w_i^* = \gamma_0 + \gamma_1 \text{centrality}_i + \gamma_2 \theta_i + \text{network\_dummy}_i + \epsilon_i, \quad (27)$$

for all the nodes of the networks we have generated.

Table 2 summarizes our estimation results of (26) and (27). As can be seen from that table, both centrality and market size positively influence a node’s equilibrium share of firms and its equilibrium wage. It is worth pointing out that the so-called ‘Home Market Effect’ (HME) – defined as a more than proportional increase in industry shares in response to an increase in local market size – always arises in both types of networks: ( $\partial \lambda_i^* / \partial \theta_i > 1$ ). This effect seems to generally hold in models without FPE (see, e.g., Takahashi, Takatsuka, and Zeng, 2013, for a discussion of the two-region case).

<sup>12</sup>Due to the high correlation between  $\lambda_i^*$  and  $n_i^*$  (see Table 1), there is no reason to look at the latter separately.

<sup>13</sup>We do not include both measures of centrality simultaneously, because of their high correlation (see Table 1).



Table 2: OLS regression results for individual nodes (Model 1).

	Dependent variable: $\lambda_i^*$					
	(i)	(ii)	(iii)	(iv)	(v)	(vi)
Closeness <sub><i>i</i></sub>	0.0671 <sup>a</sup> (24.637)		0.1019 <sup>a</sup> (23.110)		0.0347 <sup>a</sup> (11.821)	
Degree <sub><i>i</i></sub>		0.0090 <sup>a</sup> (44.085)		0.0097 <sup>a</sup> (34.725)		0.0074 <sup>a</sup> (24.247)
$\theta_i$	1.2201 <sup>a</sup> (77.116)	1.2175 <sup>a</sup> (92.053)	1.2095 <sup>a</sup> (47.772)	1.2094 <sup>a</sup> (55.957)	1.2340 <sup>a</sup> (71.466)	1.2277 <sup>a</sup> (81.959)
Constant	-0.0528 <sup>a</sup> (-26.905)	-0.0260 <sup>a</sup> (-32.700)	-0.0725 <sup>a</sup> (-24.090)	-0.0270 <sup>a</sup> (-23.765)	-0.0321 <sup>a</sup> (-15.490)	-0.0233 <sup>a</sup> (-25.825)
Network type	Both	Both	BA	BA	Equal	Equal
Network dummy	Yes	Yes	No	No	No	No
Observations	2,498	2,498	1,274	1,274	1,224	1,224
Adjusted $R^2$	0.726	0.808	0.689	0.773	0.812	0.859

	Dependent variable: $w_i^*$					
	(i)	(ii)	(iii)	(iv)	(v)	(vi)
Closeness <sub><i>i</i></sub>	0.0420 <sup>a</sup> (6.130)		0.0361 <sup>a</sup> (4.366)		0.0465 <sup>a</sup> (4.350)	
Degree <sub><i>i</i></sub>		0.0056 <sup>a</sup> (9.164)		0.0049 <sup>a</sup> (8.165)		0.0071 <sup>a</sup> (5.553)
$\theta_i$	0.1788 <sup>a</sup> (4.487)	0.1772 <sup>a</sup> (4.488)	-0.1424 <sup>a</sup> (-2.999)	-0.1428 <sup>a</sup> (-3.062)	0.4906 <sup>a</sup> (7.804)	0.4863 <sup>a</sup> (7.772)
Constant	0.9646 <sup>a</sup> (195.120)	0.9815 <sup>a</sup> (412.934)	0.9717 <sup>a</sup> (172.098)	0.9850 <sup>a</sup> (401.503)	0.9492 <sup>a</sup> (125.631)	0.9663 <sup>a</sup> (256.572)
Network type	Both	Both	BA	BA	Equal	Equal
Network dummy	Yes	Yes	No	No	No	No
Observations	2,498	2,498	1,274	1,274	1,224	1,224
Adjusted $R^2$	0.0333	0.0507	0.0200	0.0549	0.0610	0.0699

Notes: We set  $\sigma = 5$ ,  $\mu = 0.4$ , and  $\xi = 0.7$ . See Appendix D for a discussion of those choices. Simple OLS regressions. BA denotes networks generated using the Barabási and Albert (1999) algorithm.  $T$ -stats in parentheses. <sup>a</sup>, <sup>b</sup>, and <sup>c</sup> denote coefficients significant at 1%, 5%, and 10%, respectively.

Note further that both measures of centrality – closeness and degree – have a statistically strongly significant impact on the equilibrium allocation of firms across regions.<sup>14</sup> We also ran the regressions by quintiles in terms of the degree or the closeness distributions of the nodes. In both cases, the estimated coefficients for  $\theta_i$  increase monotonically with the quintiles. Thus, there is some complementarity between market size and accessibility: more accessible regions benefit more strongly from an increase in market size than more peripheral regions. In other words, increasing the size of the market in peripheral regions is unlikely to have strong impacts on the equilibrium allocation of industry.

The results pertaining to wages in the bottom panel of Table 2 deserve some comments.

<sup>14</sup>The results are identical when using the mass of firms,  $n_i^*$ , instead of the share of firms,  $\lambda_i^*$ . This finding suggests that the total mass of firms has no specific additional effect on the equilibrium allocation across regions. Since our regressions are not in logarithmic form, scaling by the total number of firms is *not* neutral. One might have expected that the total number of firms has a significant dispersive impact. Indeed, as the total number of firms rises, ‘competition’ gets tougher, and thus firms tend to disperse more.

First, as can be seen, the two measures of centrality are always positively linked to a region’s equilibrium wage. In other words, more centrally located regions or regions with better market access command higher wages, which is in line with predictions of new economic geography models and with empirical evidence (see, e.g., Mion, 2004, for Italy; and Hanson, 2005, for the us). The average equilibrium wage for ‘peripheral’ regions – in the first quintile of the degree distribution – is  $\bar{w}_{Q1} = 0.9854$ ; whereas that for more ‘central’ regions – in the fifth quintile of the degree distribution – is  $\bar{w}_{Q5} = 1.0074$ . Consequently, peripheral regions tend to specialize in the homogeneous good, paying lower wages and exporting to more central locations characterized by a high degree. The latter enjoy lower transportation costs over the network, and hence specialize in the differentiated good paying higher wages.

Second, observe that the correlation between  $\theta_i$  and  $w_i^*$  is quite low – though still positive. The intuition underlying this surprising result is as follows. Consider two regions  $i$  and  $j$ , where  $\theta_i > \theta_j$ . Assume that region  $i$  is *not fully specialized* in the production of the differentiated good, i.e., there is still some local production of the homogeneous good. If region  $i$  imports some of the homogeneous good from region  $j$ , by (11) the following condition must hold:  $p_{ji}^H = w_j \xi \tau_{ji} = p_{ii}^H = w_i \xi \tau_{ii}$ . Hence, the relative wage  $w_i/w_j$  in the two regions just depends on the relative trade costs  $\tau_{ji}/\tau_{ii}$ , but it is independent of market sizes  $\theta_i$  and  $\theta_j$ . In other words, it is just the structure of the trading network that matters, but not the distribution of market sizes. Of course, this result only holds true when a region is not fully specialized in the production of differentiated goods. Should no production of the homogeneous good take place in a region, its wage will increase with its market size – and so will the wages of the regions that export the homogeneous good to that region. We can easily confirm this conjecture by computing the correlation between  $\theta_i$  and  $w_i^*$  for the regions that do not produce any of the homogeneous good. In that case the correlation is about 0.4, instead of 0.09 when considering all regions. In other words, costly trade in the homogeneous good imposes strong conditions on wages, and those conditions partly destroy the positive link between market size and wages.

Table 3: Link between market size and equilibrium wages by node type (Model 1).

Node type	# nodes	$\bar{\theta}_i$	$\bar{w}_i^*$
Barabási and Albert			
Nodes specialized in the homogeneous good	225	0.0124	0.9974
Nodes specialized in the differentiated good	122	0.0310	1.0000
Nodes not specialized in either good	927	0.0470	0.9852
Equal probability			
Nodes specialized in the homogeneous good	205	0.0090	0.9982
Nodes specialized in the differentiated good	41	0.0240	1.0011
Nodes not specialized in either good	978	0.0460	0.9990

*Notes:* Breakdown of individual nodes by specialization type. The sample is the same than that used for the regression analysis.  $\bar{\theta}_i$  and  $\bar{w}_i^*$  denote the average market size and the average equilibrium wages of the types of nodes.

Third, centrality – both in terms of closeness *and* in terms of degree – generally has a strong impact on industry location and, to a lesser extent, on wages as explained above. A larger local market is weakly associated with higher wages, except in hub-and-spoke type BA networks (see columns (iii) and (iv) of Table 2). This latter result is surprising and requires some further explanation. As can be seen from Table 3, the largest regions are not fully specialized in the production of either type of good: their large size prevents them from being fully specialized since they cannot source all the homogeneous good that they need. Consequently, these regions have lower wages than smaller regions specialized in the differentiated good. The reason is that, as stated before, their wage is linked to the wage of the regions that supply them with the homogeneous good since they are unspecialized. In BA type networks, the largest regions have relatively low wages compared to the equal random network case, as there are on average less links with other regions in those networks. This strong non-linear effect between equilibrium wages and market size drives the negative coefficients in the lower panel of Table 2 for BA networks. Note also that the constant term in the wage regressions is close to unity, which is the theoretical value of relative wages in the absence of any differences in size and accessibility.

Our findings suggest that any analysis focusing on two regions only or disregarding the spatial structure of the trading network is likely to miss an important part of the story. It also shows that more careful theoretical analysis of multi-region trading systems is necessary, though it is well known that such an analysis is difficult to carry out in the general case when factor prices are not equalized.<sup>15</sup>

### 3.1.2 Results for the whole network

We next run regressions at the level of the network. The underlying idea is to link a measure of inequality in either the equilibrium allocation of industry or wages to measures of inequality in the distribution of market sizes and centrality in the network. We use as our inequality measure the cv of the different variables. As in the case of individual nodes, we first compute the correlations – this time across networks – for our measures of inequality. The results are reported in the bottom half of Table 4.

We then run OLS regressions to estimate the effect of the dispersion in the population shares and in centrality on the inequality in the distribution of manufacturing shares and wages. Formally, we estimate:

$$CV(\lambda_l^*) = \beta_0 + \beta_1 CV\_centrality_l + \beta_2 CV(\theta_l) + network\_dummy_l + \epsilon_l \quad (28)$$

$$CV(\mathbf{w}_l^*) = \gamma_0 + \gamma_1 CV\_centrality_l + \gamma_2 CV(\theta_l) + network\_dummy_l + \epsilon_l, \quad (29)$$

where the subscript  $l$  now denotes the network and not the individual nodes.

---

<sup>15</sup>Behrens *et al.* (2009) take an intermediate route where factor prices differ because of *exogenous* Ricardian differences. Though conceptually simpler than the case with endogenous factor prices, this approach does not allow to analyze how wages change with market size and accessibility.

Table 4: OLS regression results for the whole network (Model 1).

	Dependent variable: $CV(A_i)$					
	(i)	(ii)	(iii)	(iv)	(v)	(vi)
CV(Closeness <sub><i>t</i></sub> )	0.1530 (0.110)		-0.2892 (-0.133)		0.5292 (0.299)	
CV(Degree <sub><i>t</i></sub> )		0.2614 <sup>b</sup> (2.106)		0.1597 (1.027)		0.6380 <sup>b</sup> (2.639)
CV( $\theta_t$ )	0.6269 <sup>a</sup> (3.688)	0.7112 <sup>a</sup> (4.250)	0.7167 <sup>a</sup> (2.871)	0.7325 <sup>a</sup> (3.041)	0.5185 <sup>b</sup> (2.225)	0.8577 <sup>a</sup> (3.454)
Constant	0.4326 (1.430)	0.2515 <sup>c</sup> (1.780)	0.5920 (1.295)	0.3678 <sup>c</sup> (1.720)	0.4145 (1.056)	-0.0727 (-0.279)
Network type	Both	Both	BA	BA	Equal	Equal
Network dummy	Yes	Yes	No	No	No	No
Observations	100	100	51	51	49	49
Adjusted $R^2$	0.198	0.233	0.116	0.135	0.0655	0.187

	Dependent variable: $CV(w_t)$					
	(i)	(ii)	(iii)	(iv)	(v)	(vi)
CV(Closeness <sub><i>t</i></sub> )	0.4631 <sup>b</sup> (2.082)		0.6153 <sup>c</sup> (1.708)		0.3053 (1.157)	
CV(Degree <sub><i>t</i></sub> )		0.0154 (0.745)		0.0264 (0.997)		-0.0413 (-1.066)
CV( $\theta_t$ )	0.0284 (1.043)	0.0443 (1.588)	0.0481 (1.165)	0.0688 <sup>c</sup> (1.677)	0.0005 (0.016)	-0.0146 (-0.367)
Constant	-0.0835 <sup>c</sup> (-1.725)	-0.0011 (-0.046)	-0.1305 <sup>c</sup> (-1.726)	-0.0352 (-0.967)	-0.0323 (-0.552)	0.0697 (1.670)
Network type	Both	Both	BA	BA	Equal	Equal
Network dummy	Yes	Yes	No	No	No	No
Observations	100	100	51	51	49	49
Adjusted $R^2$	0.0476	0.0103	0.0661	0.0295	-0.0132	-0.0175

Notes: CV stands for 'coefficient of variation'. We set  $\sigma = 5$ ,  $\mu = 0.4$ , and  $\xi = 0.7$ . See Appendix D for a discussion of those choices of parameter values. Simple OLS regressions. BA denotes networks generated using the Barabási and Albert (1999) algorithm. *T*-stats in parentheses. <sup>a</sup>, <sup>b</sup>, and <sup>c</sup> denote coefficients significant at 1%, 5%, and 10%, respectively.

As can be seen from Table 4, the dispersion in market sizes,  $\theta_i$ , has a significant impact on the dispersion in the equilibrium allocation of firms, whereas the geographical structure of the trading network seems to be of lesser importance. Inequality in market size is more important for explaining inequality in the allocation of firms than the network structure. Quite surprisingly, wage inequality is not strongly linked to either inequality in the distribution of market sizes or to inequality in accessibility in the trading network. Closeness has a positive impact on wage inequality, but only in networks that have a sufficiently strong topological structure (i.e., in BA-type networks). Observe that in totally random tree networks, neither dispersion in market sizes nor in accessibility correlate significantly with dispersion in wages. One might suspect that some non-linear relationship is at work, especially since many regions can become deindustrialized, i.e., have a zero industry share (see Table 3). When a large number of regions have zero industry shares, the CV may decrease since there is no more variation coming from the deindustrialized regions. We checked formally the impact of deindustrialized regions on

equilibrium inequality. Controlling for the number of regions without industry (about 430 out of 2498, or 17.25%), we find that this variable is significant in all regressions, but that it does not change in any way the qualitative results. Thus, deindustrialized regions do not drive our key findings. They are also not driven by units of measurement issues, since we use the cv which is unit free. Last, observe that the model fit is generally much better for the dispersion of industry (top half of the table) than for the dispersion in wages (bottom half of the table). It seems thus much harder to link wage inequality to inequality in the model's fundamentals than spatial inequality in industry shares.

### 3.2 Model 2: Two differentiated sectors

We now look at the multi-region case with two differentiated CES industries. To the best of our knowledge, this has not been done until now. With two differentiated sectors, we have to examine the spatial distribution of firms in both sectors,  $\lambda_{1i}^* \equiv n_{1i}^*/(\sum_j n_{1j}^*)$  and  $\lambda_{2i}^* \equiv n_{2i}^*/(\sum_j n_{2j}^*)$ , as well as the equilibrium wages  $w_i^*$ . Simple correlations among the equilibrium values are presented in Table 5.

Table 5: Simple correlations (Model 2).

	$\lambda_{1i}^*$	$\lambda_{2i}^*$	$w_i^*$	$\theta_i$	closeness <sub>i</sub>	degree <sub>i</sub>	$\mu_{1i}$	$\mu_{2i}$
$\lambda_{1i}^*$	1							
$\lambda_{2i}^*$	-0.2194	1						
$w_i^*$	0.3202	0.3397	1					
$\theta_i$	0.6134	0.6261	0.5314	1				
closeness <sub>i</sub>	0.0008	0.0199	0.2916	0.0134	1			
degree <sub>i</sub>	0.0004	0.0182	0.3295	0.0135	0.7075	1		
$\mu_{1i}$	0.6470	-0.6587	-0.0183	-0.0157	0.0072	0.0004	1	
$\mu_{2i}$	-0.6470	0.6587	0.0183	0.0157	-0.0072	-0.0004	-1.0000	1
	CV( $A_{1i}^*$ )	CV( $A_{2i}^*$ )	CV( $w_i^*$ )	CV( $\theta_i$ )	CV(closeness <sub>i</sub> )	CV(degree <sub>i</sub> )	CV( $\mu_{1i}$ )	CV( $\mu_{2i}$ )
CV( $A_{1i}^*$ )	1							
CV( $A_{2i}^*$ )	0.2574	1						
CV( $w_i^*$ )	0.2597	0.3537	1					
CV( $\theta_i$ )	0.4656	0.6389	0.5578	1				
CV(closeness <sub>i</sub> )	0.1260	0.0786	0.0642	0.2156	1			
CV(degree <sub>i</sub> )	-0.0563	-0.1678	0.0451	-0.2254	-0.2215	1		
CV( $\mu_{1i}$ )	0.4842	-0.0632	-0.2405	-0.1219	-0.0439	-0.0404	1	
CV( $\mu_{2i}$ )	0.0456	0.5774	-0.0615	0.1009	-0.0848	-0.1315	0.0289	1

Notes: Simple correlations for 100 random tree networks with 20–30 nodes. The top panel of the table gives correlations at the level of individual nodes, whereas the bottom panel of the table gives correlations at the level of the whole network. The shares  $\lambda_{si}^*$  are computed as  $\lambda_{si}^* = n_{si}^*/(\sum_j n_{sj}^*)$ , for  $s = 1, 2$ . CV denotes the coefficient of variation, whereas  $A_1^*$ ,  $A_2^*$ ,  $w^*$ , and  $\theta$  denote the equilibrium vectors of industry shares in sectors 1 and 2, masses of firms, wages, and market size, respectively.

As can be seen from the top panel of Table 5, size and accessibility are strongly positively linked to the equilibrium industry shares and to the equilibrium wages, respectively. Although market size still positively influences wages, there is almost no correlation between our measures of centrality and the shares of firms in the two industries. As can further be seen, there

is regional specialization, as shown by the negative correlation between the equilibrium shares in both industries, as well as the positive correlation with the own expenditure share, and the negative correlation with the other industry's expenditure share. In words, this specialization is strongly driven by differences in local spending patterns, as can be seen from the last two lines of Table 5. Our finding thus extends the result on 'comparative advantage' from Proposition 1 to a multi-region setting. Note, finally, that market size has roughly the same positive impact on industry location in both industries *conditional* on expenditure shares. This is the manifestation of market size as 'absolute advantage', as subsumed by Proposition 2, which states that more centrally located regions should have, *ceteris paribus*, higher wages.

As for Model 1, we run the same regressions (26), (27), (28) and (29). The first two regressions are now run separately for the equilibrium shares of firms in each of the two sectors,  $\lambda_{1i}^*$  and  $\lambda_{2i}^*$ . In all regressions, we control for the region-specific share of expenditure on the two differentiated sectors,  $\mu_{1i}$  and  $\mu_{2i}$ .

### 3.2.1 Results for individual nodes

Table 6 shows that market size,  $\theta_i$ , and the expenditure share for the two differentiated sectors,  $\mu_{1i}$  and  $\mu_{2i}$ , are the key variables that explain the spatial distribution  $\lambda_{1i}^*$  and  $\lambda_{2i}^*$  of firms in the two sectors. The positive sign for market size is expected as labor market clearing (23) requires that the number of firms in the two sectors must sum to the population share. Once we control for local market size and the spending patterns, the centrality of a region is no longer associated with its industry share. The reason is that centrality affects both industries in the *same way*, which suggests that accessibility is akin to an absolute Ricardian advantage and should, therefore, be capitalized into factor prices.<sup>16</sup> This effect can precisely be seen from the bottom panel of Table 6. Clearly, both market size,  $\theta_i$ , and centrality are positively linked to wages,  $w_i^*$ . Regions with better access to markets and/or more trading links tend to have higher wages. Last, note that the expenditure shares  $\mu_{si}$  are nowhere near statistical significance in our wage regressions. In words, different expenditure shares affect industries differentially and, therefore, have no strong effect on regional wages. This is in line with our previous results on comparative and absolute advantage.<sup>17</sup>

---

<sup>16</sup>This result would be weakened if accessibility affected industries in different ways (as in, e.g., Hanson and Xiang, 2004). In that case, accessibility would also be in part a 'comparative advantage' and would, therefore, have a much stronger impact on industry location and not only wages.

<sup>17</sup>We performed extensive sensitivity analyses with respect to  $\sigma$  and  $\mu_{1i}$  and  $\mu_{2i}$  in Model 2. Holding the network structure constant, we study the behavior of industry location ( $\lambda_{1i}$  and  $\lambda_{2i}$ ) and regional wages,  $w_i$ , when these basic parameters change. The results consistently show that increasing  $\sigma$  in the range (1, 10] leads to higher nominal wages, while industry shares remain largely unaffected. In accord with the regression results that we report in the main text, the income shares map into the specialization. Solving a hundred times the model for each specific network and assigning random values of  $\mu_{1i}$  and  $\mu_{2i}$ , we find that industry shares present a strong correlation with income shares ( $\rho = 0.7$ ). Income shares, however, are basically uncorrelated with nominal wages.

Table 6: OLS regression results for individual nodes (Model 2).

	Dependent variable: $\lambda_{1i}^*$					
	(i)	(ii)	(iii)	(iv)	(v)	(vi)
Closeness <sub><i>i</i></sub>	-0.0034 (-1.425)		-0.0048 (-1.430)		-0.0021 (-0.605)	
Degree <sub><i>i</i></sub>		-0.0002 (-0.935)		-0.0002 (-0.854)		-0.0002 (-0.434)
$\theta_i$	0.9885 <sup>a</sup> (71.008)	0.9884 <sup>a</sup> (70.985)	1.0037 <sup>a</sup> (51.605)	1.0037 <sup>a</sup> (51.575)	0.9741 <sup>a</sup> (48.881)	0.9741 <sup>a</sup> (48.867)
$\mu_{1i}$	0.1068 <sup>a</sup> (74.763)	0.1068 <sup>a</sup> (74.737)	0.1090 <sup>a</sup> (55.252)	0.1090 <sup>a</sup> (55.219)	0.1046 <sup>a</sup> (50.567)	0.1046 <sup>a</sup> (50.548)
Constant	-0.0510 <sup>a</sup> (-27.273)	-0.0528 <sup>a</sup> (-47.617)	-0.0520 <sup>a</sup> (-20.666)	-0.0547 <sup>a</sup> (-37.983)	-0.0502 <sup>a</sup> (-19.246)	-0.0512 <sup>a</sup> (-32.047)
Network type	Both	Both	BA	BA	Equal	Equal
Network dummy	Yes	Yes	No	No	No	No
Observations	2,498	2,498	1,274	1,274	1,224	1,224
Adjusted $R^2$	0.807	0.807	0.814	0.813	0.801	0.801

	Dependent variable: $\lambda_{2i}^*$					
	(i)	(ii)	(iii)	(iv)	(v)	(vi)
Closeness <sub><i>i</i></sub>	0.0046 <sup>c</sup> (1.904)		0.0060 <sup>c</sup> (1.779)		0.0032 (0.941)	
Degree <sub><i>i</i></sub>		0.0003 (1.176)		0.0003 (1.083)		0.0002 (0.527)
$\theta_i$	0.9921 <sup>a</sup> (71.139)	0.9922 <sup>a</sup> (71.117)	0.9756 <sup>a</sup> (50.663)	0.9757 <sup>a</sup> (50.627)	1.0079 <sup>a</sup> (49.907)	1.0080 <sup>a</sup> (49.890)
$\mu_{2i}$	0.1074 <sup>a</sup> (75.016)	0.1073 <sup>a</sup> (74.970)	0.1088 <sup>a</sup> (55.664)	0.1087 <sup>a</sup> (55.615)	0.1060 <sup>a</sup> (50.564)	0.1059 <sup>a</sup> (50.533)
Constant	-0.0561 <sup>a</sup> (-29.964)	-0.0536 <sup>a</sup> (-48.900)	-0.0568 <sup>a</sup> (-22.845)	-0.0535 <sup>a</sup> (-38.461)	-0.0552 <sup>a</sup> (-20.843)	-0.0534 <sup>a</sup> (-33.548)
Network type	Both	Both	BA	BA	Equal	Equal
Network dummy	Yes	Yes	No	No	No	No
Observations	2,498	2,498	1,274	1,274	1,224	1,224
Adjusted $R^2$	0.813	0.813	0.821	0.820	0.805	0.805

	Dependent variable: $w_i^*$					
	(i)	(ii)	(iii)	(iv)	(v)	(vi)
Closeness <sub><i>i</i></sub>	0.1828 <sup>a</sup> (17.456)		0.2202 <sup>a</sup> (13.412)		0.1475 <sup>a</sup> (11.378)	
Degree <sub><i>i</i></sub>		0.0190 <sup>a</sup> (20.650)		0.0172 <sup>a</sup> (14.251)		0.0233 <sup>a</sup> (15.669)
$\theta_i$	2.0129 <sup>a</sup> (33.071)	2.0107 <sup>a</sup> (33.747)	1.8348 <sup>a</sup> (19.452)	1.8355 <sup>a</sup> (19.614)	2.1892 <sup>a</sup> (28.726)	2.1743 <sup>a</sup> (29.726)
$\mu_{1i}$	-0.0047 (-0.746)	-0.0039 (-0.638)	-0.0072 (-0.749)	-0.0074 (-0.783)	-0.0026 (-0.334)	-0.0003 (-0.038)
Constant	0.7609 <sup>a</sup> (93.127)	0.8440 <sup>a</sup> (177.839)	0.7363 <sup>a</sup> (60.309)	0.8419 <sup>a</sup> (121.672)	0.7760 <sup>a</sup> (77.848)	0.8275 <sup>a</sup> (141.247)
Network type	Both	Both	BA	BA	Equal	Equal
Network dummy	Yes	Yes	No	No	No	No
Observations	2,498	2,498	1,274	1,274	1,224	1,224
Adjusted $R^2$	0.365	0.392	0.305	0.316	0.442	0.486

Notes: We set  $\sigma = 5$ . Simple OLS regressions. BA denotes networks generated using the Barabási and Albert (1999) algorithm.  $T$ -stats in parentheses. <sup>a</sup>, <sup>b</sup>, and <sup>c</sup> denote coefficients significant at 1%, 5%, and 10%, respectively.

To summarize, regional size and expenditure patterns determine the structure of regional specialization in the two industries in Model 2, whereas accessibility has a strong impact on wages. Observe that a home market effect – defined as a more than proportional increase in industry shares in response to an increase in local market size, i.e.,  $\partial\lambda_i^*/\partial\theta_i > 1$  – generally does not arise, as shown in Table 6. The reason is that when all sectors are operating under increasing returns and face trade costs, not all of them can – by definition – exhibit home market effects (see Hanson and Xiang, 2004). In that case, an alternative definition of the HME, involving both the size  $\theta_i$  and the expenditure share  $\mu_{si}$ , would be required. To the best of our knowledge, such a definition has not been used to date in the literature.

### 3.2.2 Results for the whole network

As shown in Table 7, inequality in the distribution of market sizes and in the distribution of expenditure shares in the two differentiated sectors are the key variables that drive the inequality in the spatial distribution of firms and wages. Inequality in the network characteristics are only weakly associated with inequality in the equilibrium distributions of firms and wages. It is worth emphasizing that, as can be seen from columns (iii) and (iv) in the bottom panel of Table 7, more dispersion in the expenditure shares is negatively associated with wage inequality in the case of BA-type networks. This result is similar to the one linking the dispersion of population shares  $\theta_i$  to wage inequality in the case of BA networks in Model 1 (see the bottom panel of Table 4). In the case of equal random networks, there is no significant link between the dispersion in expenditure shares and wage dispersion.

## 3.3 Summary of results

A number of findings emerge from the foregoing analyses of the two models. Let us briefly summarize the key insights.

Starting from Model 1 with a single CES sector, we have firstly seen that accessibility has a strong impact on industry location and, to a lesser extent, on wages. This suggests that any analysis involving trade in homogeneous goods and focusing on two regions only – or disregarding the spatial structure of the trading network entirely – is likely to miss an important part of the story. Secondly, we have shown that the correlations between  $w_i^*$  and either  $\lambda_i^*$  or  $\theta_i$  are quite low, i.e., there is no strong correlation between either market size or the equilibrium industry shares and the equilibrium wages. As we have explained, this unexpected result is due to the fact that incomplete specialization in the production of the homogeneous good imposes strong restrictions on the relative wages of the trading regions, which break the link between market size and wages over the range of incomplete specialization. In that case, relative wages across regions depend on relative trade costs only but are independent of the regions' market sizes. Last, the home market effect generally holds even when trading the



Table 7: OLS regression results for the whole network (Model 2).

	Dependent variable: $CV(A_1)$					
	(i)	(ii)	(iii)	(iv)	(v)	(vi)
CV(Closeness)	0.5556 (0.711)		-0.3225 (-0.301)		1.5801 (1.356)	
CV(Degree)		0.0471 (0.660)		0.0830 (1.090)		-0.0734 (-0.422)
CV( $\theta$ )	0.6972 <sup>a</sup> (7.248)	0.7249 <sup>a</sup> (7.475)	0.7276 <sup>a</sup> (5.906)	0.7300 <sup>a</sup> (6.161)	0.6774 <sup>a</sup> (4.439)	0.6642 <sup>a</sup> (3.679)
CV( $\mu_1$ )	1.0501 <sup>a</sup> (7.761)	1.0482 <sup>a</sup> (7.748)	0.9557 <sup>a</sup> (5.610)	0.9418 <sup>a</sup> (5.604)	1.2256 <sup>a</sup> (5.518)	1.1589 <sup>a</sup> (5.112)
Constant	-0.0827 (-0.444)	-0.0060 (-0.055)	0.1517 (0.648)	0.0078 (0.059)	-0.3783 (-1.261)	0.0541 (0.227)
Network type	Both	Both	BA	BA	Equal	Equal
Network dummy	Yes	Yes	No	No	No	No
Observations	100	100	51	51	49	49
Adjusted $R^2$	0.502	0.502	0.537	0.547	0.470	0.451

	Dependent variable: $CV(A_2)$					
	(i)	(ii)	(iii)	(iv)	(v)	(vi)
CV(Closeness)	0.4631 <sup>b</sup> (2.082)		0.6153 <sup>c</sup> (1.708)		0.3053 (1.157)	
CV(Degree)		0.0303 (0.390)		-0.0135 (-0.183)		0.1105 (0.552)
CV( $\theta$ )	1.0065 <sup>a</sup> (9.735)	1.0153 <sup>a</sup> (9.816)	0.9901 <sup>a</sup> (8.663)	0.9865 <sup>a</sup> (8.792)	0.9289 <sup>a</sup> (4.660)	0.9954 <sup>a</sup> (4.459)
CV( $\mu_2$ )	1.1911 <sup>a</sup> (8.757)	1.2004 <sup>a</sup> (8.750)	1.0333 <sup>a</sup> (6.932)	1.0267 <sup>a</sup> (6.694)	1.4361 <sup>a</sup> (5.437)	1.4178 <sup>a</sup> (5.443)
Constant	-0.1996 (-1.005)	-0.2268 <sup>b</sup> (-2.027)	-0.0998 (-0.440)	-0.0889 (-0.634)	-0.3113 (-0.949)	-0.3683 (-1.649)
Network type	Both	Both	BA	BA	Equal	Equal
Network dummy	Yes	Yes	No	No	No	No
Observations	100	100	51	51	49	49
Adjusted $R^2$	0.661	0.661	0.686	0.686	0.643	0.646

	Dependent variable: $CV(w)$					
	(i)	(ii)	(iii)	(iv)	(v)	(vi)
CV(Closeness)	-0.0826 (-0.414)		-0.1392 (-0.409)		0.0639 (0.303)	
CV(Degree)		0.0154 (0.845)		0.0107 (0.437)		0.0527 <sup>c</sup> (1.757)
CV( $\theta$ )	0.1633 <sup>a</sup> (6.650)	0.1661 <sup>a</sup> (6.730)	0.1599 <sup>a</sup> (4.076)	0.1574 <sup>a</sup> (4.126)	0.1697 <sup>a</sup> (6.143)	0.1993 <sup>a</sup> (6.409)
CV( $\mu_1$ )	-0.0688 <sup>b</sup> (-1.992)	-0.0680 <sup>c</sup> (-1.974)	-0.1209 <sup>b</sup> (-2.230)	-0.1240 <sup>b</sup> (-2.292)	0.0078 (0.195)	0.0180 (0.462)
Constant	0.0270 (0.569)	-0.0028 (-0.102)	0.0726 (0.974)	0.0352 (0.826)	-0.0449 (-0.827)	-0.0864 <sup>b</sup> (-2.110)
Network type	Both	Both	BA	BA	Equal	Equal
Network dummy	Yes	Yes	No	No	No	No
Observations	100	100	51	51	49	49
Adjusted $R^2$	0.339	0.342	0.301	0.301	0.433	0.468

Notes: CV stands for ‘coefficient of variation’. We set  $\sigma = 5$ . Simple OLS regressions. BA denotes networks generated using the Barabási and Albert (1999) algorithm.  $T$ -stats in parentheses. <sup>a</sup>, <sup>b</sup>, and <sup>c</sup> denote coefficients significant at 1%, 5%, and 10%, respectively.

homogeneous good is costly, provided that it is less costly than trading the differentiated good.

Turning next to Model 2 with two CES sectors, both absolute market size – as captured by  $\theta_i$  – and centrality – as measured by either closeness or the degree distribution – are capitalized into factor prices, thus showing that they constitute absolute advantage affecting all industries in the same way. Differences in spending patterns – as captured by the  $\mu_{si}$  – are however capitalized into industry structure, thus showing that they constitute comparative advantage affecting industries differently. Our findings, therefore, extend the theoretical results of Behrens and Ottaviano (2011), which have been derived with two regions only, to a multi-region setting.

Last, it is worth pointing out that the effects of accessibility and market size on wages are an order of magnitude larger in Model 2 than in Model 1 (compare Tables 6 and 2). As we have explained, the reason is that the equalization of prices in the traded homogeneous sector imposes strong restrictions on the determination of wages among trading partners when specialization is incomplete (a very frequent case). This in turn breaks the link between accessibility and market size in the wage determination. In a nutshell, market size and centrality matter all the more the more industries are subject to trade costs and increasing returns to scale.

## 4 Numerical application to Spanish regions

While the foregoing numerical simulations highlight regularities of our multi-region trade models without FPE, they provide no sense of how well those models perform when confronted with data. The aim of this section is hence to use calibrated versions of the models to check their fit with the data and to run a series of counterfactuals. To this end, we compute the equilibria of the two models using Spanish provincial data in two years: 1980 and 2007 (see Appendix D for a description of the data). This is an interesting period because it coincides with significant changes in demographic trends, and with important infrastructure improvements. Paluzie *et al.* (2007) discuss the migration trends in Spain from rural to urban areas that, starting in the sixties, still characterized demographic trends in the eighties. The fundamental tendency was the agglomeration of population in ever larger urban areas. Zofío *et al.* (2014) show that the decentralization of public administration – as Spain joined the European Community – accompanied by substantial funding from the European Regional Development Plan (ERDP) helped to finance remarkable improvements in the road network. Along with price changes in the transportation sector, mainly driven by salaries and fuel, generalized transport costs fell by about 15% over the period we consider. In a nutshell, our study period was one of important changes in the population distribution and in transport costs, both of which should have a strong influence on the spatial equilibrium structure of the economy.

Our aim in the remainder of this section is twofold. First, we compare the equilibrium distribution of economic activity predicted by our models with the data. Doing so will allow us to assess to what extent the models can ‘replicate’ the observed distributions. Second, we use

the model to run some simple counterfactuals with respect to changes in demographic trends and transportation costs. We disentangle the role of market size from the role of transportation costs by shutting down one of the two channels when running our counterfactuals. More precisely, we first look at the equilibria of the models in the absence of any changes in the labor force between 1980 and 2007, i.e., when changes are ‘solely’ driven by changes in transportation costs. Second, we repeat the exercise by assuming that there are no changes in transportation costs between 1980 and 2007, so that changes are ‘solely’ driven by changes in the spatial distribution and in the size of the labor force.

#### 4.1 Equilibrium distributions vs. observed distributions

The equilibrium distributions of firms in 1980 and 2007, as well as the equilibrium wages, are summarized in Table 8. Our results show large disparities in the distribution of firms across provinces, and those disparities increased between 1980 and 2007. In each year, the distribution of firms varies from almost 0% to about 16%–25%. Not surprisingly, the provinces of Madrid and Barcelona have the highest shares of firms in both models. These provinces are the largest – in terms of population shares – which, as reported in the previous sections, is the main determinant of firm shares (followed, to a lesser extent, by centrality that benefits Madrid as the geographical center of the Spanish infrastructure network). On the contrary, very small provinces situated in the Iberian Peninsula plateau (plain or meseta) are almost devoid of production (e.g., the provinces surrounding Madrid such as Toledo, Cuenca, Guadalajara, Segovia, or Ávila have a really negligible share of firms).

To tentatively gauge the predictive power of the models, we check the statistical significance of the differences between the observed distributions of production in ‘differentiated products’ and those associated with the equilibria of the models:  $\lambda_i^*$  (Model 1) and  $\lambda_{1i}^*$  and  $\lambda_{2i}^*$  (Model 2). Besides Pearson’s  $r$  and Spearman’s  $\rho$  coefficients of correlation, we also test the equality of distributions by way of a Kolmogorov-Smirnoff test. Table 9 reports large and significant correlations, both for linear (Pearson) and rank (Spearman) dependencies. The Pearson standard correlation ranges from 0.8638 in Model 1 for 2007 to a remarkable 0.9910 for Model 2 in the same year. The maximum values for the Spearman correlations correspond to the same models and year. Additionally, the hypothesis of equality of distributions cannot be generally rejected, except for the 2007 distribution in Model 1. Our results show that solving the models using real data yields model equilibrium distributions of economic activity that are in many cases statistically hard to distinguish from those observed in the real economy.

Turning to wages, we however do not find large correlations between those proxied by GDP per employee (our empirical counterpart for ‘wages’ at the aggregate level) and the solutions to the two models. Hence, while the models perform well in terms of their spatial predictions of economic activity, they perform much worse in terms of their predictions for prices. There

Table 8: Simulation results for Model 1 and Model 2.

Region	Model 1				Model 2					
	$\lambda_i^*$		$w_i^*$		$\lambda_{1i}^*$		$\lambda_{2i}^*$		$w_i^*$	
	1980	2007	1980	2007	1980	2007	1980	2007	1980	2007
Almeria	0.012	0.020	1	1	0.010	0.015	0.012	0.016	1	1
Cadiz	0.02	0.009	1.046	1.027	0.021	0.022	0.024	0.025	1.122	1.051
Cordoba	0.015	0.004	1.073	1.049	0.017	0.014	0.019	0.015	1.089	1.031
Granada	0.010	0	1.052	1.033	0.017	0.015	0.019	0.017	0.937	0.964
Huelva	0.005	0.001	1.061	1.043	0.009	0.009	0.010	0.010	1.074	1.002
Jaen	0.014	0.002	1.067	1.045	0.014	0.010	0.016	0.012	0.988	0.957
Malaga	0.024	0.028	1.040	1.033	0.023	0.029	0.025	0.032	1.051	1.042
Sevilla	0.058	0.096	1.077	1.067	0.033	0.036	0.037	0.040	1.260	1.169
Huesca	0.001	0	1.073	1.047	0.007	0.006	0.006	0.005	0.931	0.892
Teruel	0	0	1.089	1.059	0.005	0.004	0.004	0.003	0.953	0.933
Zaragoza	0.027	0.015	1.067	1.034	0.026	0.027	0.023	0.023	1.012	0.954
Asturias	0.031	0.012	1.020	1.008	0.035	0.023	0.034	0.022	0.878	0.812
Cantabria	0.011	0.012	1.049	1.046	0.016	0.013	0.016	0.014	0.919	0.934
Avila	0	0	1.102	1.076	0.005	0.004	0.005	0.004	1.034	1.089
Burgos	0.004	0.003	1.077	1.048	0.011	0.009	0.011	0.010	1.003	1.016
Leon	0.015	0	1.062	1.047	0.017	0.010	0.018	0.010	0.98	0.946
Palencia	0.005	0.001	1.106	1.087	0.005	0.004	0.006	0.004	1.071	1.045
Salamanca	0.012	0	1.087	1.052	0.010	0.007	0.010	0.007	1.101	1.020
Segovia	0	0	1.109	1.097	0.004	0.004	0.005	0.005	1.097	1.159
Soria	0.002	0.003	1.117	1.098	0.003	0.002	0.003	0.003	1.080	1.089
Valladolid	0.023	0.033	1.116	1.08	0.013	0.013	0.013	0.013	1.175	1.145
Zamora	0.002	0	1.082	1.061	0.007	0.004	0.007	0.004	1.002	0.981
Albacete	0.004	0	1.092	1.062	0.010	0.009	0.009	0.008	1.047	1.048
Ciudad Real	0.008	0.001	1.085	1.050	0.013	0.011	0.012	0.010	1.020	1.045
Cuenca	0	0	1.100	1.073	0.006	0.005	0.006	0.004	1.007	1.066
Guadalajara	0	0	1.102	1.079	0.004	0.005	0.004	0.005	1.059	1.109
Toledo	0	0	1.081	1.066	0.014	0.014	0.013	0.014	1.090	1.147
Barcelona	0.171	0.190	1.047	0.991	0.14	0.145	0.133	0.138	1.271	1.095
Girona	0	0	1.035	1.009	0.018	0.02	0.017	0.019	0.884	0.789
Lleida	0	0	1.053	1.035	0.012	0.012	0.012	0.012	0.882	0.888
Tarragona	0.014	0.005	1.088	1.061	0.017	0.019	0.016	0.018	1.153	1.062
Alicante	0.045	0.057	1.092	1.065	0.032	0.038	0.033	0.039	1.282	1.178
Castellon	0.016	0.014	1.093	1.068	0.014	0.013	0.014	0.015	1.213	1.105
Valencia	0.064	0.067	1.034	1.011	0.058	0.057	0.060	0.061	1.16	1.103
Badajoz	0.014	0.007	1.061	1.038	0.014	0.012	0.017	0.015	0.997	0.99
Caceres	0.006	0	1.066	1.041	0.010	0.006	0.012	0.009	0.931	0.945
A coruna	0.036	0.021	1.036	1.02	0.033	0.024	0.034	0.025	1.112	0.984
Lugo	0.012	0	1.047	1.036	0.016	0.007	0.017	0.008	0.989	0.916
Orense	0.016	0.003	1.054	1.052	0.016	0.006	0.017	0.007	1.072	1.010
Pontevedra	0.035	0.038	1.051	1.046	0.030	0.021	0.031	0.022	1.184	1.081
Madrid	0.157	0.256	1.142	1.095	0.123	0.181	0.113	0.165	1.505	1.475
Murcia	0.017	0.02	1.055	1.065	0.022	0.028	0.025	0.033	1.086	1.117
Navarra	0.006	0.004	1.056	1.033	0.016	0.017	0.015	0.018	0.978	0.959
Alava	0.012	0.022	1.123	1.090	0.009	0.009	0.008	0.008	1.302	1.156
Guipuzcoa	0.025	0.031	1.053	1.102	0.022	0.021	0.019	0.018	1.28	1.204
Vizcaya	0.031	0.01	1.041	1.003	0.033	0.031	0.029	0.026	1.074	0.936
La Rioja	0.02	0.016	1.139	1.028	0.008	0.008	0.007	0.008	1.283	1.026
Mean	0.021	0.021	1.072	1.05	0.021	0.021	0.021	0.021	1.077	1.035
Std. Dev	0.039	0.054	0.030	0.029	0.030	0.038	0.028	0.035	0.133	0.117
Max.	0.171	0.256	1.142	1.102	0.140	0.181	0.133	0.165	1.505	1.475
Min.	0	0	1	0.991	0.003	0.002	0.003	0.003	0.878	0.789

Notes: Simulations use the following values. For Model 1, we let  $\sigma = 5$ ,  $\mu = 0.4$ , and  $\xi = 0.7$ . For Model 2, we let  $\sigma = 5$ . See Appendix D for a discussion of these choices.

are two possible reasons for this. First, GDP per capita – though widely used in the literature (see Head and Mayer, 2004) – is only a crude proxy for wages. Second, as shown in the previous section, the multi-region simulated models do not deliver clear results as to the roles of market size and centrality on wages. It is thus not surprising that their empirical fit to wage data is also fairly weak.

Table 9: Differences between observed and model distributions.

Test		Pearson's <sup>1</sup> $r$		Spearman's <sup>2</sup> $\rho$		Kolmogorov-Smirnov <sup>3</sup>	
Model	Share	1980	2007	1980	2007	1980	2007
1	$\lambda_i^*$	0.9386 (0.000)	0.8638 (0.000)	0.7222 (0.000)	0.7261 (0.000)	0.1915 (0.3207)	0.4043 (0.0006)
2	$\lambda_{1i}^*$	0.9627 (0.000)	0.9334 (0.000)	0.8321 (0.000)	0.8582 (0.000)	0.1702 (0.4662)	0.2128 (0.2096)
	$\lambda_{2i}^*$	0.9401 (0.000)	0.991 (0.000)	0.9339 (0.000)	0.9886 (0.000)	0.2128 (0.2096)	0.1277 (0.8117)

Notes: <sup>1,2</sup>The null hypothesis is that both variables are independent; <sup>3</sup>The null hypothesis is that both variables come from the same continuous distribution.  $p$ -values for all tests in parenthesis.

## 4.2 Population, transport costs, and trends in inequality

The equilibria computed in the foregoing section can be used to analyze to what extent the models capture the process of agglomeration that has taken place in Spain between 1980 and 2007, and which resulted in a more unequal distribution of manufacturing activity.

Figure 1: Manufacturing distributions for Model 1.

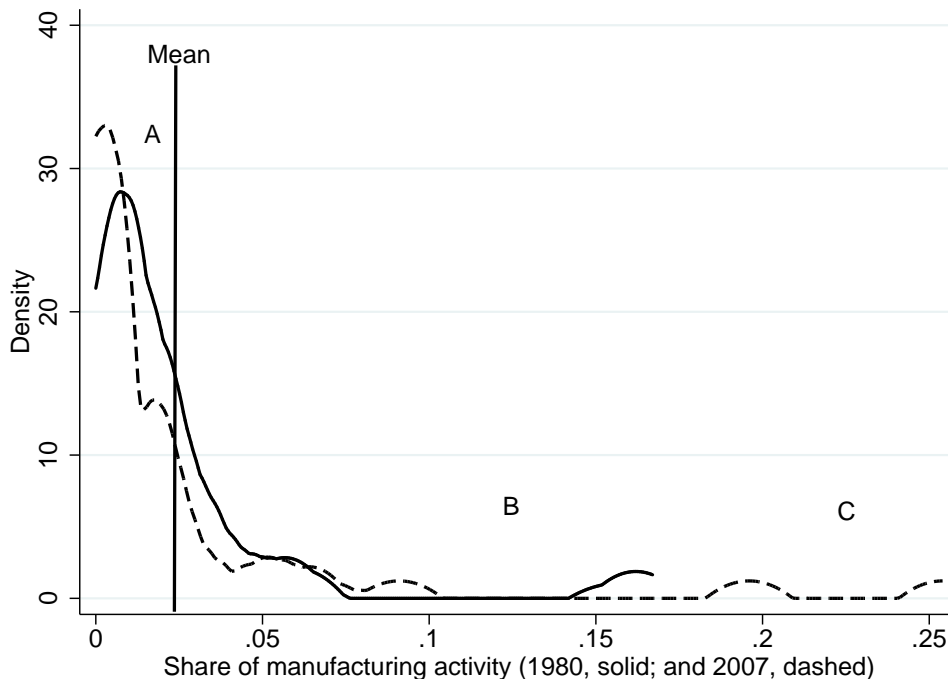


Figure 1 depicts the changes in the equilibrium manufacturing shares for Model 1. The equilibrium distributions in 1980 and 2007 are displayed as solid and as dashed lines, respectively. As can be seen from Figure 1, although the distributions are fairly similar (because there is a lot of inertia in spatial structures), the one in 2007 exhibits a higher density for low values of the manufacturing shares in the area identified by A (just below the mean value of 0.021, depicted by the vertical line). Another difference pointing towards an increase in inequality is the agglomeration of manufacturing activity that can be seen in 2007, with the equilibrium shares of Madrid and Barcelona driving this process. Indeed, both provinces increased their values from 0.167 to 0.196 and from 0.157 to 0.254, respectively. This evolution is visible from the dilation of the right tail of the distribution in the area identified by C. For values in the range between 0.05 and 0.15, both distributions display similar densities (area B) between the two years.<sup>18</sup>

To provide a quantitative sense of the increase in inequality, we have computed the Gini indices  $G$  for the distribution of observed manufacturing shares and for the equilibria of the two models in both years. Both models capture the increase in inequality, even if *both clearly overstate it*. Observed inequality in the distribution of the manufacturing sector increases by 0.60% (from  $G^{80} = 0.7700$  to  $G^{07} = 0.7816$ ), while Model 1 yields an increase in inequality of 12.84% (from  $G_{M1}^{80} = 0.7805$  to  $G_{M1}^{07} = 0.8808$ ). Using the equilibria from Model 2, the observed increase in inequality is 2.12% for the manufacturing sector (from  $G_1^{80} = 0.7609$  to  $G_1^{07} = 0.7770$ ), and 2.41% (from  $G_2^{80} = 0.7762$  to  $G_2^{07} = 0.7950$ ) for the service sector, respectively.<sup>19</sup> The model again overpredicts these values at 11.69% ( $G_{M2,1}^{80} = 0.6841$  to  $G_{M2,1}^{07} = 0.7641$ ), and 11.90% ( $G_{M2,2}^{80} = 0.6683$  to  $G_{M2,2}^{07} = 0.7479$ ), respectively.

We may thus conclude that while the model reasonably well predicts the spatial distribution of manufacturing in Spain for a given year, it overpredicts the impacts of changes in population or changes in transportation costs on that spatial distribution.

### 4.3 Counterfactuals

Keeping in mind the caveat from the previous section, we finally run two counterfactuals, the aim of which is to simulate the spatial equilibrium that would prevail if only population changes to its 2007 values, but not the transportation costs which are kept fixed at their 1980 values, and vice versa. Put differently, in the first counterfactual we fix the transport costs to their 1980 values and use observed population changes; whereas in the second counterfactual, we fix population to their 1980 values and use observed changes in transportation costs. In so doing, we can compare the ‘pure’ effect of population changes conditional on transport costs,

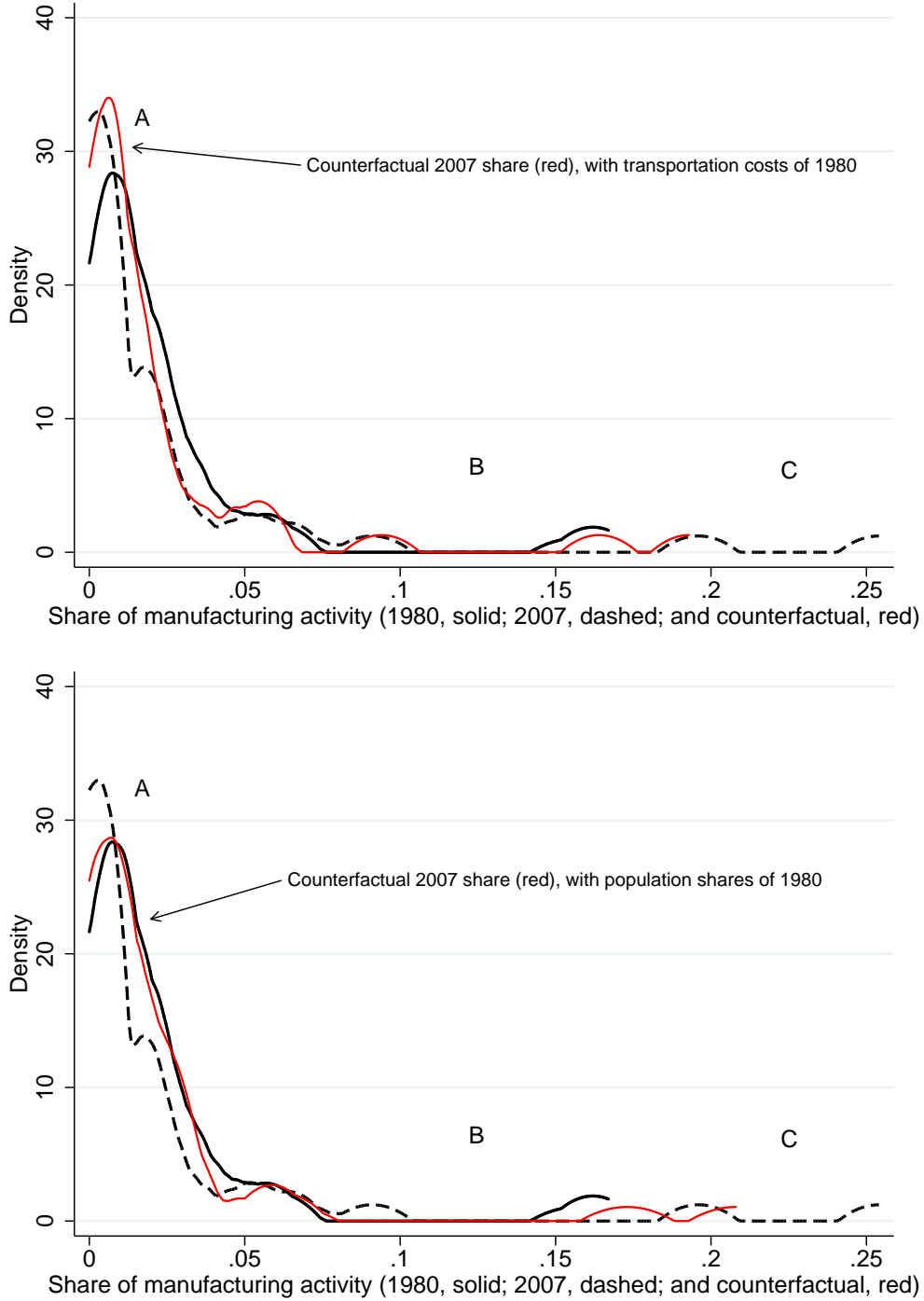
---

<sup>18</sup>Similar results are observed for the two distributions of shares in Model 2. The results are available from the authors upon request.

<sup>19</sup>In Appendix D, we explain that the sectors are not defined in the same way in the two models. Hence the difference in the changes in observed inequality in the data.

and the 'pure' effect of changes in transport costs conditional on population. We compare the equilibria of the model in 1980 and 2007 to those derived in the counterfactuals to determine how each change contributes to the overall shift in the manufacturing shares.<sup>20</sup>

Figure 2: Manufacturing distributions for Model 1 (with counterfactual distribution).



<sup>20</sup>Note that this is of course not an exact decomposition since the observed changes are not the sum of the changes in the two alternative scenarios.

The top panel of Figure 2 depicts the distributions of the counterfactual industry shares (in red) for the case where only population changes; whereas the bottom panel of the figure depicts the same change for the case where only transport costs change. Table 10 summarizes the detailed results, with the first superscript referring to the reference year for the population share,  $\theta_i$ , and the second one referring to the bilateral transportation costs,  $\phi_{ij}$ .

As can be seen from the top panel of Figure 2, the increase in the density of the left tail of the distribution of manufacturing shares is mainly driven by the change in the geographical distribution of the labor force rather than the reduction in transportation costs. As can be seen from the bottom panel of Figure 2, the effect of the latter is rather small, despite the fact that over the 1980–2007 period the fall in the average value of the Generalized Transport Costs was large.<sup>21</sup> As this change was similar across provinces, their relative position in the network remained basically unchanged. Observe that the fall in transport costs has slowed down the process of agglomeration, as can be seen from the bottom panel of Figure 2.

Generally, when comparing the counterfactual distributions with the observed ones in 1980 and 2007, Figure 2 reveals that changes in the spatial distribution of regional market shares have a stronger predictive power of changes in industrial specialization than changes in transportation costs. Our results thus suggest that the reallocation of the labor force was the main driver of agglomeration and larger inequalities as reflected by the change in the Gini indices. Note that these results are compatible with those obtained in the simulations presented in Section 3, and particularly with those in Table 2 for Model 1. In that case, the equilibrium shares  $\lambda_i^*$  depend mainly on the  $\theta_i$  rather than on network features as captured by transportation costs (i.e., closeness or degree). As the relative position of the provinces in the trading network did not change much between 1980 and 2007, this explains the stability in the distributions of the spatial equilibria when considering changes in this variable only.

To conclude on a policy note, observe that after three decades of significant investments in the road network, the distribution of industry shares had not changed much in Spain. Thus, these investments do not seem to have contributed much to territorial cohesion – though the main goal of infrastructure investment in the eyes of policy makers is often to ‘reduce regional inequality’. In fact, the opposite occurred: Madrid and Barcelona had larger shares of economic activity in 2007 than in 1980. These changes in industry shares were mostly driven by population reshuffling, and little by decreasing transportation costs. The financial efforts of transport improvements did apparently not translate into higher cohesion and lower inequality.<sup>22</sup>

---

<sup>21</sup>The fall in  $GTC_{ij}$  amounted to 14.14%. This fall corresponds to an increase of 109.65% in the average  $\phi_{ij}$ , thus implying that the freeness of trade more than doubled.

<sup>22</sup>One word of caution is in order. Our approach does not capture the fact that the population change between 1980 and 2007 would possibly have been different in the absence of changes in transportation costs. Conversely, the changes in transportation costs would possibly have been different between 1980 and 2007 in the absence of population movements. Hence, changing one parameter while holding fixed the other is only a partial exercise (though, we believe, a suggestive one).



Table 10: Counterfactual results for Models 1 and 2.

Region	Model 1				Model 2					
	$\lambda_i^{+80,07}$	$\lambda_i^{+07,80}$	$w_i^{+80,07}$	$w_i^{+07,80}$	$\lambda_{1i}^{+80,07}$	$\lambda_{1i}^{+07,80}$	$\lambda_{2i}^{+80,07}$	$\lambda_{2i}^{+07,80}$	$w_i^{+80,07}$	$w_i^{+07,80}$
Almeria	0.007	0.018	1	1	0.010	0.015	0.012	0.016	1	1
Cadiz	0.008	0.021	1.005	1.061	0.021	0.022	0.024	0.025	1.080	1.088
Cordoba	0.014	0.008	1.007	1.114	0.017	0.014	0.019	0.015	1.082	1.029
Granada	0.009	0.007	1.006	1.086	0.017	0.015	0.019	0.017	0.997	0.898
Huelva	0	0.005	1.006	1.077	0.009	0.009	0.010	0.010	1.037	1.032
Jaen	0.008	0.009	1.007	1.095	0.014	0.010	0.017	0.012	1.011	0.924
Malaga	0.021	0.035	1.005	1.049	0.023	0.029	0.025	0.032	1.056	1.037
Sevilla	0.094	0.06	1.007	1.112	0.033	0.036	0.037	0.040	1.203	1.223
Huesca	0	0	1.007	1.104	0.007	0.006	0.006	0.005	0.927	0.881
Teruel	0	0	1.009	1.116	0.005	0.004	0.004	0.003	0.975	0.897
Zaragoza	0.014	0.028	1.007	1.101	0.026	0.026	0.023	0.023	0.986	0.973
Asturias	0.017	0.021	1.004	1.032	0.035	0.023	0.033	0.022	0.878	0.807
Cantabria	0.01	0.01	1.006	1.062	0.016	0.014	0.016	0.013	0.980	0.868
Avila	0	0	1.011	1.110	0.005	0.004	0.005	0.004	1.136	0.969
Burgos	0.008	0.003	1.009	1.104	0.011	0.009	0.011	0.009	1.064	0.947
Leon	0	0.007	1.007	1.083	0.017	0.010	0.018	0.010	1.026	0.893
Palencia	0.023	0.003	1.011	1.150	0.005	0.004	0.006	0.004	1.103	1.003
Salamanca	0.001	0.008	1.009	1.095	0.010	0.007	0.010	0.007	1.088	1.02
Segovia	0.001	0	1.012	1.097	0.004	0.004	0.005	0.004	1.187	1.054
Soria	0.004	0.002	1.012	1.143	0.003	0.002	0.003	0.002	1.131	1.021
Valladolid	0.036	0.022	1.011	1.150	0.013	0.013	0.013	0.013	1.193	1.125
Zamora	0.004	0	1.009	1.101	0.007	0.004	0.007	0.004	1.057	0.914
Albacete	0	0.002	1.009	1.122	0.010	0.009	0.009	0.008	1.089	0.996
Ciudad Real	0.004	0.007	1.009	1.114	0.013	0.011	0.012	0.010	1.089	0.969
Cuenca	0	0	1.011	1.113	0.006	0.005	0.006	0.004	1.106	0.951
Guadalajara	0	0	1.011	1.079	0.004	0.005	0.005	0.005	1.124	1.032
Toledo	0	0	1.010	1.059	0.013	0.014	0.014	0.014	1.168	1.059
Barcelona	0.164	0.173	1.002	1.134	0.140	0.145	0.133	0.138	1.125	1.227
Girona	0.001	0	1.004	1.024	0.018	0.020	0.017	0.019	0.806	0.859
Lleida	0	0	1.006	1.061	0.012	0.012	0.012	0.012	0.916	0.847
Tarragona	0.013	0.019	1.007	1.110	0.017	0.019	0.016	0.018	1.087	1.120
Alicante	0.056	0.051	1.008	1.133	0.033	0.038	0.033	0.039	1.196	1.256
Castellon	0.019	0.015	1.008	1.126	0.013	0.014	0.014	0.014	1.139	1.168
Valencia	0.053	0.065	1.006	1.053	0.057	0.058	0.060	0.060	1.137	1.118
Badajoz	0.016	0.012	1.007	1.08	0.014	0.012	0.017	0.015	1.038	0.945
Caceres	0	0.004	1.008	1.076	0.009	0.007	0.012	0.009	1.001	0.869
A coruna	0.018	0.027	1.004	1.044	0.033	0.024	0.034	0.025	1.056	1.033
Lugo	0.011	0.003	1.006	1.066	0.016	0.007	0.017	0.007	1.019	0.875
Orense	0.017	0.006	1.007	1.076	0.016	0.006	0.017	0.007	1.126	0.939
Pontevedra	0.054	0.026	1.006	1.052	0.030	0.021	0.031	0.022	1.157	1.098
Madrid	0.193	0.208	1.009	1.253	0.124	0.18	0.113	0.166	1.464	1.509
Murcia	0.015	0.024	1.007	1.084	0.021	0.028	0.025	0.032	1.126	1.075
Navarra	0.004	0.006	1.007	1.089	0.015	0.018	0.016	0.018	0.983	0.950
Alava	0.03	0.011	1.011	1.200	0.009	0.009	0.008	0.008	1.193	1.251
Guipuzcoa	0.038	0.026	1.008	1.033	0.022	0.021	0.019	0.019	1.249	1.225
Vizcaya	0.003	0.028	1.005	1.069	0.034	0.031	0.029	0.027	0.974	1.023
La Rioja	0.009	0.02	1.009	1.193	0.008	0.008	0.007	0.008	1.059	1.233
Mean	0.021	0.021	1.008	1.095	0.021	0.021	0.021	0.021	1.077	1.026
Std. Dev	0.043	0.045	0.003	0.049	0.03	0.037	0.028	0.035	0.108	0.144
Max.	0.193	0.208	1.012	1.253	0.140	0.180	0.133	0.166	1.464	1.509
Min.	0	0	1	1	0.003	0.002	0.003	0.002	0.806	0.807

Notes: The simulations use the following values. For Model 1, we set  $\sigma = 5$ ,  $\mu = 0.4$ , and  $\xi = 0.7$ . For Model 2, we let  $\sigma = 5$ . See Appendix D for a discussion of those choices.

## 5 Conclusions

We have investigated the geographical distribution of industries and wages in a asymmetric multi-region models without factor price equalization. Using systematic numerical simulations for two different trade models – one with a homogeneous and a differentiated sector, and another with two differentiated sectors – we have studied whether and how size and accessibility are linked to the equilibrium industry shares and to wages.

Our key findings can be summarized as follows. First, absolute local market size and accessibility are crucial in explaining a region's wage. This is due to the fact that absolute market size – as measured by the population size of a region – and accessibility – as measured by network centrality or the degree distribution of a region – affect all industries in similar ways, i.e., constitute a region's *absolute advantage*. This effect is stronger and more systematic in models where all sectors are subject to transport costs and exhibit increasing returns to scale.

Second the relative local market size of industries – as captured by their expenditure shares – is crucial in explaining a region's industrial composition. This is due to the fact that relative spending patterns do not affect all industries in the same way, i.e., constitute a region-specific *comparative advantage*. In a nutshell – and very much in line with Ricardian trade theory – absolute advantage translates into higher wages, whereas comparative advantage maps into specialization patterns.

Third, the correlation between equilibrium wages and equilibrium industry shares is rather low in both models, thus suggesting that the two adjustment channels work largely independently. Empirical tests and formal definitions of the home market effect should take into account both dimensions – industry location and wages – in order to be relevant. To the best of our knowledge, tests looking simultaneously at industry location and factor prices have not yet been devised.

Finally, when applying the two models to Spanish data – using Generalized Transport Costs between regions as a measure of trade frictions – we find that the models generally predict well the distribution of industries, yet predict less well the spatial patterns in wages. The latter may be due to the fact that GDP per capita – though often used in the literature – is a rather crude proxy for wages. It may, however, also be linked to the fact that regional differences in accessibility are generally less pronounced than regional differences in population shares. Thus the second effect may dwarf the former in the applications.

**Acknowledgements.** We thank our discussants, Ted Rosenbaum and Jacques Thisse, as well as Rafael Moner-Colonques, Inmaculada C. Álvarez, and conference participants at the 2014 NARSC Meetings in Atlanta, the 3rd Workshop on Urban Economics in Barcelona, and the 10th Meetings on Economic Integration in Castellón for very helpful comments and suggestions. Barbero acknowledges financial support from the Spanish Ministry of Education (AP2010-1401).

Behrens acknowledges financial support from the CRC Program of the Social Sciences and Humanities Research Council (SSHRC) of Canada for the funding of the *Canada Research Chair in Regional Impacts of Globalization*. Barbero and Zofío acknowledge financial support from the Spanish Ministry of Science and Innovation (ECO2010-21643 and ECO2013-46980-P). Part of the paper was written while Barbero was visiting UQÀM, the hospitality of which is gratefully acknowledged. The views expressed in this paper, as well as all remaining errors, are ours.

## References

- [1] Baldwin, Richard, Rikard Forslid, Philippe Martin, Gianmarco I.P. Ottaviano, and Frédéric Robert-Nicoud. 2003. *Economic Geography and Public Policy*. Princeton, NJ: Princeton University Press.
- [2] Barabási, Albert-László, and Réka Albert. 1999. "Emergence of scaling in random networks." *Science* 286(5439): 509–512.
- [3] Behrens, Kristian, and Gianmarco I.P. Ottaviano. 2011. "General equilibrium trade theory and firm behavior." In: Bernhofen, Daniel, David Greenaway, Rod Falvey, and Udo Kreike-meier (eds.) *Palgrave Handbook of International Trade*. Basingstoke, UK: Palgrave MacMillan, pp. 119–159.
- [4] Behrens, Kristian, Andrea R. Lamorgese, Gianmarco I.P. Ottaviano, and Takatoshi Tabuchi. 2007. "Changes in transport and non-transport costs: local vs global impacts in a spatial network." *Regional Science and Urban Economics* 37(6): 625–648.
- [5] Behrens, Kristian, Andrea R. Lamorgese, Gianmarco I.P. Ottaviano, and Takatoshi Tabuchi. 2009. "Beyond the Home Market Effect: Market size and specialization in a multi-country world." *Journal of International Economics* 79(2): 259–265.
- [6] Behrens, Kristian, and Pierre M. Picard. 2007. "Welfare, home market effects, and horizontal foreign direct investment." *Canadian Journal of Economics* 40(4): 1118–1148.
- [7] Behrens, Kristian, and Pierre M. Picard. 2011. "Transportation, freight rates, and economic geography." *Journal of International Economics* 85(2): 280–291.
- [8] Broda, Christian, and David E. Weinstein. 2006. "Globalization and the gains from variety." *Quarterly Journal of Economics* 121(2): 541–585.
- [9] Davis, Donald R. 1998. "The home market, trade, and industrial structure." *American Economic Review* 88(5): 1264–1276.

- [10] Davis, Donald R., and David E. Weinstein. 2003. "Market access, economic geography and comparative advantage: an empirical test." *Journal of International Economics* 59(1): 1–23.
- [11] Duranton, Gilles, and Matthew A. Turner. 2012. "Urban growth and transportation." *Review of Economic Studies* 79(4): 1407–1440.
- [12] FBBVA (Several years), Renta Nacional de España y su Distribución Provincial, Bilbao: Fundación BBVA.
- [13] Freeman, Linton C. 1979. "Centrality in social networks: Conceptual clarification." *Social Networks* 1(3): 215–239.
- [14] Gómez-Antonio, Miguel, and Bernard Fingleton. 2012. "Analyzing the impact of public capital stock using the NEG wage equation: A spatial panel data approach." *Journal of Regional Science* 47(2): 486–502.
- [15] Hanson, Gordon H. 2005. "Market potential, increasing returns and geographic concentration." *Journal of International Economics* 67(1): 1–24.
- [16] Hanson, Gordon H., and Chong Xiang. 2004. "The Home-Market Effect and bilateral trade patterns." *American Economic Review* 94(4): 1108–1129.
- [17] Head, Keith, and Thierry Mayer. 2004. "The empirics of agglomeration and trade." In: Henderson, J. Vernon, and Jacques-François Thisse (eds.) *Handbook of Regional and Urban Economics*, vol. 4. Amsterdam: North Holland, pp. 2609–2669.
- [18] Head, Keith, Thierry Mayer, and John Ries. 2002. "On the pervasiveness of Home Market Effects." *Economica* 69(275): 371–390.
- [19] Helpman, Elhanan, and Paul R. Krugman. 1985. *Market Structure and Foreign Trade*. Cambridge, MA: MIT Press.
- [20] Krugman, Paul R. 1980. "Scale economies, product differentiation and the pattern of trade." *American Economic Review* 70(5): 950–959.
- [21] Krugman, Paul R. 1991. "Increasing returns and economic geography." *Journal of Political Economy* 99(3): 483–499.
- [22] MFOM. 2007a. "Encuesta Permanente de Transporte de Mercancías por Carretera. Año 2007." Madrid: Ministerio de Fomento.
- [23] MFOM. 2007b. "Observatorio de Costes de Transporte de Mercancías por Carretera. Año 2007." Madrid: Ministerio de Fomento.

- [24] Mion, Giordano. 2004. "Spatial externalities and empirical analysis: The case of Italy." *Journal of Urban Economics* 56(1): 97–118.
- [25] Ottaviano, Gianmarco I.P., Takatoshi Tabuchi, and Jacques-François Thisse. 2002. "Agglomeration and trade revisited." *International Economic Review* 43((2)): 409–436.
- [26] Ottaviano, Gianmarco I.P., and Jacques-François Thisse. 2004. "Agglomeration and economic geography." In: J.V. Henderson and J.-F. Thisse (eds.) *Handbook of Regional and Urban Economics*, vol. 4, Elsevier, North-Holland, pp. 2563–2608.
- [27] Paluzie, Elisenda, Jordi Pons, Javier Silvestre, and Daniel Tirado. 2009. "Migrants and market potential in Spain over the twentieth century: A test of the new economic geography." *Spanish Economic Review* 11(4): 243–265.
- [28] Picard, Pierre M., and Dao-Zhi Zeng. 2005. "Agricultural sector and industrial agglomeration." *Journal of Development Economics* 77(1): 75–106.
- [29] Pons, Jordi, Elisenda Paluzie, Javier Silvestre, and Daniel A. Tirado. 2007. "Testing the New Economic Geography: Migrations and industrial agglomerations in Spain." *Journal of Regional Science* 47(2): 289–313.
- [30] Rauch, James E. 1999. "Networks versus markets in international trade." *Journal of International Economics* 48(1): 7–35.
- [31] Takahashi, Toshiaki, Hajime Takatsuka, and Dao-Zhi Zeng. 2013. "Spatial inequality, globalization, and footloose capital." *Economic Theory* 53(1): 213–238.
- [32] Takatsuka, Hajime, and Dao-Zhi Zeng. 2012. "Trade liberalization and welfare: Differentiated-good versus homogeneous-good markets." *Journal of the Japanese and International Economies* 26(3): 308–325.
- [33] Xie, Feng, and David Levinson. 2008. "Topological evolution of surface transportation networks." *Computers, Environment, and Urban Systems* 33(3): 211–223.
- [34] Yu, Zhihao. 2005. "Trade, market size, and industrial structure: revisiting the home-market effect." *Canadian Journal of Economics* 38(1): 255–272.
- [35] Zeng, Dao-Zhi, and Toru Kikuchi. 2009. "Home Market Effect and trade costs." *Japanese Economic Review* 60(2): 253–270.
- [36] Zofío, José L., Ana M. Condeço-Melhorado, Andrés Maroto-Sánchez, and Javier Gutiérrez. 2014. "Generalized transport costs and index numbers: A geographical analysis of economic and infrastructure fundamentals." *Transportation Research Part A: Policy and Practice* 67(C): 141–157.

# Appendix

## A. Factor price equalization

Assume that the homogeneous good can be costlessly traded across all regions. This is the case usually considered in the literature (e.g., Helpman and Krugman, 1985). Marginal cost pricing then implies that the price of the homogeneous good is equal to the wage, which must be the same everywhere. In other words, factor price equalization (FPE) holds.

In a multi-region world, the assumption of FPE has a major technical drawback. To see this, ask under what conditions FPE will hold? Clearly, FPE will hold if and only if some homogeneous good is produced in every region. Following Behrens, Lamorgese, Ottaviano, and Tabuchi (2007), a *sufficient condition* is that

$$\theta_i > \mu, \quad \forall i = 1, 2, \dots, M. \quad (\text{A-1})$$

When (A-1) holds for all regions, and when trade in the homogeneous good is free, we have  $w_i = 1$  for all  $i = 1, 2, \dots, M$ . Observe that condition (A-1) is extremely restrictive. Consider, e.g., a world with 30 regions. If market sizes  $\theta_i$  were identical across regions, we must have  $\mu < 1/30$ . This is already very restrictive. But in our case, since we randomly assign the shares  $\theta_i$  to regions, we may have very small shares in some cases. In those cases, the foregoing restriction can never be met for ‘reasonable values’ of  $\mu$ .

Although condition (A-1) is technically speaking only a *sufficient* condition – i.e., we may still have FPE even when it is violated – it seems still very unlikely to be met in general. Another potential problem in the FPE version of the model is that it displays a much larger share of ‘corner equilibria’, i.e., equilibria in which some regions are deindustrialized and do not host any of the differentiated sector. We have simulated the model with FPE and find that the number of nodes with a zero industry share is 920 out of 2498, i.e. 36.82%. This is a large number, so that regression methods dealing with zeros may be required to analyze the general properties of these equilibria.

In a nutshell, the FPE model does not make much sense in a world with many regions, neither theoretically nor empirically, and it is difficult to implement consistently for reasonable values of  $\mu$ . We thus disregard it in the remainder of this paper.

## B. Generating random tree networks

We use two different algorithms for generating random tree networks. The first one is based on Barabási and Albert (1999). This algorithm starts with a network having  $M_0$  linked nodes. Then, it adds new nodes one by one, up to  $M_T$  nodes in total, where  $M_T$  is the number of nodes of the network (i.e., the number of regions in the model). Each time a new node is added to the network at iteration  $t$ , it is connected to  $M_{t-1}$  pre-existing nodes. The probability

of being linked to an existing node during iteration  $t$  depends on the degree of that node in the following way:  $p_{it} = \text{deg}(i_{t-1}) / [\sum_j \text{deg}(j_{t-1})]$ , where  $p_{it}$  is the probability of being linked to node  $i$  at iteration  $t$ , and where  $\text{deg}(i_{t-1})$  is the degree of node  $i$  at iteration  $t - 1$ . The Barabási and Albert (1999) preferential attachment algorithm tends to create networks with some nodes that have a high degree, who are very well connected, and other nodes with a very low degree, who are badly connected. Put differently, the resulting network tends to have hub-and-spoke characteristics. By setting the initial number of nodes to  $M_0 = 2$ , and by setting the number of links for new nodes to  $m = 1$ , we ensure that the resulting network is a connected tree with  $M_T - 1$  links.

In the second algorithm we use, new nodes are added to preexisting nodes with equal attachment probability, which means that the probability of being linked to node  $i$  at iteration  $t$  does not depend on the degree of node  $i$ . Formally, we have  $p_{it} = 1/M_{t-1}$ , where  $M_{t-1}$  is the number of nodes in the network when adding the new node at iteration  $t$ .

Observe that the average degree of the tree network is equal to  $2(M_T - 1)/M_T$ , independently of the algorithm used to generate it. The reason is that in an undirected graph, the degree sum formula is  $\sum_j \text{deg}(j) = 2|E|$ , where  $|E|$  is the number of links in the network. Since in the generated tree networks there are  $M_T - 1$  links, the degree sum formula becomes  $2(M_T - 1)$ . Then, the average degree of the network, defined as the degree sum over the number of nodes in the network, is equal to  $2(M_T - 1)/M_T$ .

Observe further that the standard deviation of the degree of the nodes in the network will usually be higher in networks using the Barabási and Albert (1999) algorithm than in totally random tree networks. The reason is that this algorithm tends to generate a few nodes with a high degree, and a lot of nodes with a very low degree.

Last, when generating random links in the networks, we assume that the freeness of trade,  $\phi_{ij}$ , between adjacent nodes  $i$  and  $j$  is given by  $1/5$ . Hence, the freeness of trade between two nodes  $i$  and  $k$ , linked by a path  $\mathcal{P} = \{i, j_1, j_2, \dots, j_{n-1}, k\}$  of length  $n$ , is given by

$$\phi_{ik} = \prod_{(j,l) \in \mathcal{P}} \phi_{jl}. \tag{B-1}$$

We use only shortest paths in the network, which are computed using the Floyd-Warshall algorithm. Because we work with trees, the shortest path is uniquely determined.

### C. Details on the numerical implementation

We first use the algorithms described in Appendix B to generate random networks. In all cases, we compute the equilibria of the two models for the *same* set of networks. Hence, the results are directly comparable across models. For computational reasons, we generate random networks with between 20 and 30 nodes, the number of nodes being itself random (and drawn

from a uniform distribution). Larger networks require too long to solve in the case with a homogeneous good.

To solve the model, we transform the spatial equilibrium conditions (16) into complementary slackness conditions as follows:

$$[\text{RMP}_i(\mathbf{n}) - 1] n_i = 0, \quad i = 1, 2, \dots, M, \quad (\text{C-1})$$

where we make explicit the dependence of the real market potential on the whole distribution of firms  $\mathbf{n} = (n_1, n_2, \dots, n_M)$ .

**Model 1: One differentiated sector and one homogeneous sector.** We add as nonlinear inequality constraints the equilibrium conditions (13) in the homogeneous good market, the labor market clearing conditions (14), and the complementary slackness conditions (15) for exports of the homogeneous good:

$$\begin{aligned} \frac{(1 - \mu) w_i \theta_i}{\min_k \{w_k \xi \tau_{ki}\}} - (\tilde{X}_{ii} + \sum_{j \neq i} \tilde{X}_{ji}) &= 0, \quad \forall i \\ \theta - n_i \mu - (\tau_{ii} \tilde{X}_{ii} + \sum_{j \neq i} \xi \tau_{ij} \tilde{X}_{ij}) &= 0, \quad \forall i \\ \tilde{X}_{ij} \left[ w_i \xi \tau_{ij} - \min_k \{w_k \xi \tau_{kj}\} \right] &= 0, \quad \forall i \end{aligned}$$

Furthermore, the following bounds for the variables are imposed:  $w_i > 0$  for all  $i$  and  $\tilde{X}_{ji} \geq 0$  for all  $i$  and  $j$ . We also have the constraints that  $n_i \geq 0$  for all  $i$ . Note that the presence of the min function, which is not differentiable, makes it more difficult to solve the problem. To overcome this problem, we replace all occurrences of the min function with a new variable,  $z_i$ . To make sure that this new variable  $z_i$  will be equal to the minimum, we subtract it from the objective function (i.e., it works as a penalty). Thus, the solver will maximize it. We add the constraint that it should not exceed the delivered price of the homogeneous good:  $z_i \leq w_j \xi \tau_{ji}, \forall i, j$ . In doing so, we make sure that – in the final iteration –  $z_i$  is equal to the minimum delivered price of the good.

We transform (C-1) into an equivalent problem that consists in minimizing the sum of squared residuals subject to the set of equilibrium constraints. The numerical implementation



of the minimization problem – when substituting out the min operator – is as follows:

$$(\mathcal{P}_1) \left\{ \begin{array}{l} \min_{\mathbf{n}} \sum_{i=1}^M \{[\text{RMP}_i(\mathbf{n}) - 1] n_i\}^2 - \sum_{i=1}^M z_i \\ \text{RMP}_i(\mathbf{n}) \leq 1, \quad i = 1, 2, \dots, M \\ \frac{(1 - \mu) w_i \theta_i}{z_i} - (\tilde{X}_{ii} + \sum_{j \neq i} \tilde{X}_{ji}) = 0, \quad \forall i \\ \theta_i - n_i \mu - (\tau_{ii} \tilde{X}_{ii} + \sum_{j \neq i} \xi \tau_{ij} \tilde{X}_{ij}) = 0, \quad \forall i \\ z_i \leq w_j \xi \tau_{ji}, \quad \forall i, j \\ \theta_i - n_i \mu \geq 0, \quad \forall i \\ \tilde{X}_{ij} [w_i \xi \tau_{ij} - z_j] = 0, \quad \forall i \\ n_i \geq 0, \quad \forall i, \quad w_i > 0, \quad \forall i \\ \tilde{X}_{ji} \geq 0, \quad \forall i, j \end{array} \right. \quad (\text{C-2})$$

As starting values for the solver, we use the population share,  $\theta_i$ , for the mass of firms, i.e.,  $n_i^0 = \theta_i$ . For the wages, we use  $w_i^0 = 1$  for all  $i$ . Last, we start with zeros for trade in the homogeneous good,  $\tilde{X}_{ki}^0 = 0$ , and  $\tilde{X}_{ii}^0 = (1 - \mu)\theta_i$  for the domestic supply of the homogeneous good to the local market.

**Model 2: Two differentiated sectors.** For the model with two differentiated sectors, we minimize the sum of the squared residuals of the two complementary slackness conditions of the real market potential for each sector:

$$[\text{RMP}_{si}(\mathbf{n}_s, \mathbf{w}) - 1] n_{si} = 0, \quad i = 1, 2, \dots, M, \quad s = 1, 2.$$

The minimization problem is similar to the one in the case with a homogeneous good, but with two real market potential functions with the number of firms,  $n_{si}$ , in each sector, the inclusion of the wages, and the constraint on the number of firms and the population shares:

$$(\mathcal{P}_2) \left\{ \begin{array}{l} \min_{\mathbf{n}_1, \mathbf{n}_2, \mathbf{w}} \sum_{i=1}^M \{[\text{RMP}_{1i}(\mathbf{n}_1, \mathbf{w}) - 1] n_{1i}\}^2 + \sum_{i=1}^M \{[\text{RMP}_{2i}(\mathbf{n}_2, \mathbf{w}) - 1] n_{2i}\}^2 \\ \text{RMP}_{1i}(\mathbf{n}_1, \mathbf{w}) \leq 1, \quad \forall i \\ \text{RMP}_{2i}(\mathbf{n}_2, \mathbf{w}) \leq 1, \quad \forall i \\ \theta_i = n_{1i} + n_{2i}, \quad \forall i \\ n_{1i} > 0, \quad \forall i, \quad n_{2i} > 0, \quad \forall i, \quad w_i > 0, \quad \forall i \end{array} \right. \quad (\text{C-3})$$

We solve the problems  $(\mathcal{P}_1)$  and  $(\mathcal{P}_2)$  for their equilibria  $\{n_i^*, w_i^*\}$ , and  $\{n_{1i}^*, n_{2i}^*, w_i^*\}$ , respectively. We use the MATLAB function `fmincon` with the *interior-point* algorithm. The code is available upon request.

## D. Data and calibration

We work with Spanish provincial data at the NUTS-3 level, totaling 47 observations.<sup>23</sup> Table 11 provides details on the variables needed to solve the different models. For Model 1, these include the labor force shares ( $\theta_i$ ), the gross value added shares in the differentiated sector (the observed  $n_i$  or  $\lambda_i$ ), and the mean of the bilateral transportation costs ( $\tau_{ij}$ ). Population and industrial gross value added – our proxy for the differentiated production in the economy – for 1980 are obtained from the ‘Spanish Domestic Income and its Distribution by Provinces’ (FBBVA) publication. The 2007 data come from the Spanish National Statistics Institute (Instituto Nacional de Estadística, INE). The FBBVA data on private gross value added at the provincial level is disaggregated into Agriculture, Energy, Industry, Construction, and Services. Bilateral shipping costs are measured as the monetary value of the generalized transportation cost ( $GTC_{ij}$ ) of delivering one ton of cargo between origin  $i$  and destination  $j$ . Zofío *et al.* (2014) describe the model assuming a cost minimizing behavior on the part of transportation firms, and determine the least cost optimal itineraries using geographical information systems that account for the actual road network in those years. In Table 11, we provide the mean value of all bilateral transportation costs for each province, i.e.,  $\overline{GTC}_{ij}^t = \frac{1}{47} \sum_{i=1}^{47} GTC_{ij}^t$ . Following the definition of the freeness of trade,  $\phi_{ij}$ , transport costs are computed as follows:

$$\phi_{ij} = \tau_{ij}^{1-\sigma} = \left( \frac{GTC_{ij}}{\min\{GTC_{ij}\}} \right)^{1-\sigma} \in [0, 1]. \quad (\text{D-1})$$

As for the structural parameters  $\mu$  and  $\sigma$  of the model, few studies have attempted to test the main propositions of new trade theory and new economic geography using Spanish data. Pons *et al.* (2007) estimate a migration equation based on an NEG model, and obtain a value for  $\sigma$  between 2.8 and 4.2, conditional on the values of the other parameters. Gómez-Antonio and Fingleton (2012) adopt a value of  $\sigma = 6.25$  when analyzing the impact of the public capital stock on Spanish productivity. Their choice is justified on the grounds that it coincides with the key estimates in the literature (e.g., Table 5 in Head and Mayer, 2004). More recently, Broda and Weinstein (2006) estimate the elasticities of substitution for traded goods imports to the US using SITC rev2 for 1972–1988, and SITC rev3 for 1990–2001 at the 3-, 4-, and 5-digit levels, respectively. At the 3-digit level and across all goods, they find a mean elasticity of 6.8 from 1972–1988 and of 4.0 from 1990–2001, respectively. Looking only at differentiated goods – as defined using the Rauch (1999) classification – at the 4-digit level, they find a mean elasticity of 5.2 from 1972–1988 and of 4.7 from 1990–2001, respectively. Since the estimates obtained by these authors are probably the best currently available, and since they are roughly in line with the estimates obtained for Spain, we take the midpoint value of  $\sigma = 5$  (as we also assumed in the numerical simulations performed in the previous sections). Turning to

---

<sup>23</sup>We use all Spanish provinces of the Iberian Peninsula (i.e., we exclude the Balearic islands and the Canary islands) because our measures of transport costs are derived from road freight transportation.

the expenditure share on the differentiated product,  $\mu$ , we use the expenditure shares for manufacturing goods in total domestic demand coming from the household budget survey published by INE, which in 2007 was 41.92% (data for 1980 is unfortunately unavailable, but this share exhibits remarkable stability both in time and across developed countries, fluctuating around this value depending on the economic cycle). For simplicity, we round the value to  $\mu = 0.4$  (as we also assumed in the numerical simulations).

Since wages are endogenous, we require additional data to test whether the results of the calibrated model match the observed values. In particular, we need information on wages. The latter are obtained, as in many previous studies, by dividing aggregate GDP by the labor force (see the literature review in Head and Mayer, 2004). For Model 1, we associate the homogeneous sector with agriculture, while the differentiated sector corresponds to the manufacturing industry. As for the parameter  $\xi$  capturing the relative level of trade cost of the homogeneous good compared to the differentiated good, we adopt a value 0.7. Based on data from the ‘Ongoing Survey on Freight Road Transportation’, carried out by the Ministry of Transport (see Ministerio de Fomento, MFOM, 2007a), we can calculate a comparative range of relative freight costs in terms of tons-kilometer.<sup>24</sup> The difference in the cost of shipping homogeneous and differentiated products ranges from 0.7 to 1, with an average around 0.8. To keep consistency with the values adopted in the previous section, we take the lower bound for  $\xi$ .

Finally, besides regional labor shares and bilateral trade costs, we need to identify two differentiated sectors for Model 2. We associate the first differentiated sector with manufacturing plus energy, whereas services are associated with the second sector. We leave out agriculture – which is more homogeneous – and construction – which is essentially non-tradable – from the analysis. We determine expenditure shares to match the production side from the expenditure household survey, with the first share corresponding to manufacturing and utilities (processed food, clothing, water, electricity, . . .) and the second one to services (health, communication, leisure, education, accommodation, . . .). These shares are, unfortunately, only available at the NUTS-2 regional level (States or Comunidades Autónomas), and they are an average of all NUTS-3 provinces included in each region. As a result we apply the regional values to all provinces of a region. Although this reduces the regional variation, it is the only way we can use that required piece of information. Table 11 below summarizes the data that we use.

---

<sup>24</sup>The ‘Ongoing Survey on Freight Road Transportation’ classifies shipments of manufactured goods according to Council Regulation (EC) No 1172/98, and the prevalent type of vehicle used to transport each type of good, along with the information provided by the Observatory of Road Freight Transport Costs on each type of vehicle (MFOM, 2007b).

Table 11: Data for the 47 peninsular Spanish provinces (NUTS-3 level).

Model	Data 1980					Data 2007					Model 2
	All	All	Model 1	Model 2	Model 2	All	All	Model 1	Model 2	Model 2	
Region	Labor %	Mean GTC €	G.V.A. Industry %	G.V.A. Ind+Ene %	G.V.A. Services %	Labor %	Mean GTC €	G.V.A. Industry %	G.V.A. Ind+Ene %	G.V.A. Services %	$\mu_1$ Ind+Ene
Almeria	0.011	902.224	0.004	0.004	0.008	0.016	754.879	0.004	0.004	0.013	0.435
Cadiz	0.023	958.657	0.019	0.020	0.023	0.024	829.657	0.015	0.018	0.023	0.435
Cordoba	0.019	731.552	0.011	0.011	0.014	0.015	637.700	0.010	0.010	0.013	0.435
Granada	0.018	837.438	0.007	0.007	0.017	0.017	705.907	0.007	0.006	0.016	0.435
Huelva	0.010	889.677	0.012	0.013	0.008	0.009	751.991	0.007	0.009	0.009	0.435
Jaen	0.016	691.018	0.010	0.010	0.012	0.011	575.984	0.008	0.007	0.010	0.435
Malaga	0.024	959.042	0.010	0.010	0.028	0.031	821.833	0.008	0.008	0.032	0.435
Sevilla	0.036	799.573	0.023	0.022	0.036	0.038	685.078	0.026	0.031	0.035	0.435
Huesca	0.007	665.786	0.006	0.007	0.005	0.006	626.479	0.005	0.005	0.005	0.488
Teruel	0.005	562.000	0.005	0.007	0.003	0.004	517.986	0.003	0.005	0.003	0.488
Zaragoza	0.024	582.48	0.022	0.021	0.024	0.025	538.244	0.035	0.033	0.023	0.488
Asturias	0.034	936.268	0.033	0.042	0.018	0.023	810.024	0.027	0.029	0.022	0.471
Cantabria	0.016	778.184	0.017	0.016	0.014	0.014	649.923	0.016	0.015	0.013	0.471
Avila	0.005	457.276	0.002	0.002	0.004	0.004	403.746	0.002	0.003	0.003	0.460
Burgos	0.011	550.503	0.012	0.012	0.009	0.009	468.094	0.017	0.016	0.008	0.460
Leon	0.017	642.546	0.011	0.015	0.009	0.010	544.999	0.007	0.012	0.010	0.460
Palencia	0.006	536.972	0.005	0.006	0.004	0.004	445.716	0.005	0.005	0.003	0.460
Salamanca	0.010	557.103	0.006	0.008	0.007	0.007	472.840	0.005	0.005	0.007	0.460
Segovia	0.005	443.097	0.003	0.003	0.004	0.004	384.724	0.002	0.002	0.004	0.460
Soria	0.003	458.082	0.002	0.002	0.003	0.002	385.037	0.003	0.002	0.002	0.460
Valladolid	0.013	487.398	0.017	0.015	0.011	0.013	412.244	0.015	0.014	0.013	0.460
Zamora	0.007	597.455	0.003	0.004	0.004	0.004	493.528	0.002	0.002	0.004	0.460
Albacete	0.010	560.532	0.006	0.005	0.007	0.008	486.351	0.005	0.005	0.006	0.471
Ciudad Real	0.012	532.628	0.009	0.012	0.008	0.011	450.946	0.008	0.010	0.008	0.471
Cuenca	0.006	466.723	0.003	0.003	0.005	0.005	405.691	0.002	0.002	0.004	0.471
Guadalajara	0.004	451.418	0.005	0.005	0.004	0.005	394.952	0.006	0.006	0.004	0.471
Toledo	0.014	471.100	0.010	0.010	0.009	0.014	410.415	0.015	0.014	0.01	0.471
Barcelona	0.137	940.605	0.200	0.188	0.148	0.142	879.725	0.211	0.197	0.151	0.477
Girona	0.017	1166.402	0.017	0.016	0.018	0.020	1049.276	0.019	0.017	0.020	0.477
Lleida	0.012	797.744	0.011	0.012	0.011	0.012	720.807	0.010	0.009	0.011	0.477
Tarragona	0.017	747.118	0.024	0.030	0.015	0.019	667.449	0.021	0.022	0.019	0.477
Alicante	0.033	689.785	0.036	0.033	0.033	0.039	585.947	0.028	0.027	0.038	0.454
Castellon	0.014	633.634	0.016	0.017	0.012	0.014	562.700	0.024	0.023	0.012	0.454
Valencia	0.059	631.054	0.059	0.055	0.06	0.059	552.074	0.057	0.054	0.056	0.454
Badajoz	0.016	687.265	0.005	0.005	0.011	0.013	564.444	0.005	0.006	0.010	0.419
Caceres	0.011	602.853	0.005	0.007	0.006	0.008	502.318	0.002	0.004	0.007	0.419
A coruna	0.034	1084.937	0.022	0.025	0.025	0.025	861.847	0.018	0.025	0.023	0.452
Lugo	0.017	957.134	0.006	0.007	0.008	0.007	768.696	0.007	0.006	0.006	0.452
Orense	0.016	851.836	0.006	0.007	0.006	0.007	660.066	0.005	0.006	0.005	0.452
Pontevedra	0.031	1084.473	0.017	0.016	0.019	0.022	827.120	0.025	0.022	0.018	0.452
Madrid	0.118	442.748	0.130	0.124	0.217	0.173	387.143	0.128	0.136	0.215	0.483
Murcia	0.024	769.986	0.023	0.026	0.028	0.030	615.794	0.023	0.024	0.026	0.435
Navarra	0.016	625.343	0.025	0.023	0.016	0.018	575.838	0.031	0.029	0.016	0.471
Alava	0.009	639.970	0.02	0.018	0.008	0.009	556.701	0.024	0.021	0.009	0.494
Guipuzcoa	0.020	785.055	0.041	0.037	0.021	0.020	659.040	0.046	0.041	0.019	0.494
Vizcaya	0.031	777.486	0.059	0.057	0.036	0.029	682.024	0.042	0.047	0.032	0.494
La Rioja	0.008	490.447	0.008	0.007	0.006	0.008	513.364	0.012	0.011	0.007	0.482
Mean	0.021	700.268	0.021	0.021	0.021	0.021	601.220	0.021	0.021	0.021	0.461
Std. Dev	0.029	189.805	0.04	0.037	0.043	0.036	156.292	0.041	0.039	0.043	0.018
Max.	0.137	1166.402	0.200	0.188	0.217	0.173	1049.276	0.211	0.197	0.215	0.494
Min.	0.003	442.748	0.002	0.002	0.003	0.002	384.724	0.002	0.002	0.002	0.419

Notes: For Model 2, we have  $\mu_2 = 1 - \mu_1$  by definition.