REM WORKING PAPER SERIES

# Market potential, road accessibility, and firm births: evidence from twenty years of road investment

Mauricio S. de Carvalho, Patrícia C. Melo, Bruno T. Rocha, Isabel Proença, João de Abreu e Silva

REM Working Paper 0364-2025

January 2025

**REM – Research in Economics and Mathematics**
Rua Miguel Lúpi 20,
1249-078 Lisboa,
Portugal

ISSN 2184-108X

**REM – Research in Economics and Mathematics**

Rua Miguel Lupi, 20
1249-078 LISBOA
Portugal

Telephone: +351 - 213 925 912
E-mail: rem@iseg.ulisboa.pt

https://rem.rc.iseg.ulisboa.pt/

**Market potential, road accessibility, and firm births: evidence from twenty years of road investment**

Mauricio S. de Carvalho[1], Patrícia C. Melo[1], Bruno T. Rocha[1], Isabel Proença[2], João de Abreu e Silva[3]

[1] REM-UECE & ISEG/Universidade de Lisboa

[2] REM-CEMAPRE & ISEG/Universidade de Lisboa

[3] CERIS & IST/Universidade de Lisboa

Acknowledgments:

This work was supported by Fundação para a Ciência e Tecnologia (FCT) through grants PTDC/GES-GCE/4046/2021 (TASSEL- Growth or Relocation? Transport Accessibility and Economic Activity Location), DOI 10.54499/2021.04760.BD (Carvalho), UIDB/05069/2020 (REM: Melo/UECE and Proença/CEMAPRE), and 2023.07270.CEECIND/CP2889/CT0001 (Rocha/UECE).

Address for correspondence:

mauriciodesc@phd.iseg.ulisboa.pt (preferred); pmelo@iseg.ulisboa.pt;

brunorocha@iseg.ulisboa.pt; isabelp@iseg.ulisboa.pt;

jabreu@tecnico.ulisboa.pt

**Market potential, road accessibility, and firm births: evidence from twenty years of road investment**


## ABSTRACT

This paper investigates the causal effects of road accessibility, measured by market potential, on firm births in Portuguese municipalities between 1991 and 2016, a period marked by significant road improvements. We address the endogeneity of market potential by employing instrumental variables within Poisson Pseudo-maximum Likelihood estimates with fixed effects, which we refer to as "non-local time-variant historical instruments". Our estimated elasticities for firm births range from 1.6 to 1.9 for the 1-year interval and 1.2 to 1.3 for the 5-year interval. Additionally, we find a greater positive effect on firm births when excluding the metropolitan areas of Lisbon and Porto, which is indicative evidence of a heterogeneous spatial effect. We also find that the impact of the enlarged market potential is heterogeneous across sectors. Our results are robust to changes in model specification and the usage of alternative measures of the instruments.



Keywords: road investment; firm births; market potential; Poisson Pseudo-maximum Likelihood; instrumental variables

JEL: O18; R39; R49

## 1. Introduction

The spatial distribution of economic activity and its determinants has been a major topic of academic research since Von Thünen (1842), Marshall (1890), and Weber (1909). The topic continues to be active, and during the last decades, research on industrial location decisions has intensified (Arauzo-Carod et al., 2010; Balbontin & Hensher, 2019; Carpenter et al., 2021). Alongside advances in econometric modeling, spatial economic theory, and microdata availability, we can ascribe two policy-related reasons for the growing interest in the determinants of firm location. First, understanding the factors that influence firms' location decisions can better guide policymakers in developing more efficient and effective policies, which, in turn, enhance a region's attractiveness, foster job creation, and drive economic growth. Second, given that many governments use transport investment to promote regional growth and cohesion, it is important to investigate if the empirical evidence supports this claim.

Portugal provides a relevant case study due to its considerable investment in road infrastructure. This investment was possible due to the country's accession to the European Union (EU) in 1986, making it eligible for structural funds aimed at increasing economic development in the EU's poorest regions. Total road investment (including public, private, and European contributions) increased from 0.74% of GDP in the 1980s to 1.52% in the 2000s. In particular, total motorway investment increased from 0.07% of GDP in the 1980s to 0.59% in the 2000s (Pereira & Pereira, 2018). According to Eurostat, Portugal's motorway network expanded from 196 km in 1986 to 2,737 km in 2011, reaching its peak in 2014 with 3,065 km. Consequently, Portugal has the third largest motorway network per capita in 2021, after Croatia and Spain.[1]

Despite the substantial investment in motorways, there is relatively little empirical evidence for Portugal on the effect of improvements in road accessibility on firm location (Holl, 2004b, 2004c; Melo et al., 2010), contrasting with countries like Spain, which also invested heavily in motorways.[2] One of the main contributions of this study is to provide novel evidence on the impact of improved road accessibility on firm births, using market

[1] Authors calculations based on data for motorways (Eurostat/motorways) and population (Eurostat/population).

[2] To cite a few, see Alama-Sabáter et al. (2011), Alañón-Pardo and Arauzo-Carod (2013), Alañón-Pardo et al. (2018), Arauzo-Carod and Manjón-Antolín (2012), Holl (2004a), Holl and Mariotti (2018a, 2018b), Jofre-Monseny et al. (2014) and Manjón-Antolín and Arauzo-Carod (2011).

potential as the central measure. While several other studies examine the influence of transport accessibility on firm location decisions (e.g., Coughlin & Segev, 2000; Ghani et al., 2016; Mejia-Dorantes et al., 2012; Nilsson & Smirnova, 2016; Percoco, 2016), only a handful of studies have employed market potential as the variable of interest for capturing this effect (Gibbons et al., 2019; Holl, 2004b, 2004c; Holl & Mariotti, 2018b; Melo et al., 2010; Otsuka, 2008; Rupasingha & Marré, 2020).[3]

Specifically, our study contributes to the literature in the following ways. First, we address endogeneity bias from simultaneity between firm location and market potential using instrumental variables (IV) in the context of Poisson Pseudo-maximum Likelihood (PPML) with fixed effects (PPML-FE), following Lin and Wooldridge (2019). To our knowledge, this is the first study to combine IV and PPML-FE to examine firm location determinants. We propose a "non-local time-variant historical instrument" that interacts a time-variant component with historical data. The former is based on the road accessibility of the "non-local" region relative to the national level. The "non-local" region is defined as the NUTS2 region of the municipality, excluding both the municipality itself and its neighbors.[4] The historical variables include the inverse of the mean straight-line distance to the nearest Roman major roads and the 18th-century itineraries computed for the municipalities within that non-local region. The total length of these historical roads is also considered as an alternative measure. By combining non-locality with historical information, we aim to provide a composite IV (CIV) that can be used with fixed effects (FE) and is more likely to be exogenous. To implement this approach, we rely on the PPML method, popularized by Santos Silva and Tenreyro (2006, 2011), which has only been explored by a few studies in the context of firm location (Coll-Martínez et al., 2022; Jofre-Monseny et al., 2014; Moeller, 2018; Schlegel & Backes-Gellner, 2023).

Second, in contrast to most existing studies, we measure road accessibility using road-based travel times instead of simple measures of physical straight-line distances.[5] Third, we provide insights into the heterogeneous effects of market potential on firm

[3] In our study, we use establishment data to compute the count of births aggregated by municipality and sector; in the text, we use the terms “firms,” “plants,” and “establishments” interchangeably.
[4] Municipalities must share a border and be in the same NUTS2 region to be classified as neighbors.
[5] Travel times distances were obtained from the longitudinal spatial road database constructed as part of the project TiTuSS. See Afonso et al. (2025) for details on the construction of the database and Afonso et al. (2024) to access the database.

location by examining its impact across manufacturing and services, with the findings showing that the effect of market potential is concentrated in the services sector.

Finally, concerning Portugal specifically, we have two additional contributions. First, we address a gap in the literature by providing causal evidence on the impact of improved road accessibility on firm location in Portugal, as research in this area remains sparse compared to other countries. To date, only three studies – Holl (2004b, 2004c) and Melo et al. (2010) – have explored the relationship between road infrastructure and spatial patterns of firm birth in Portugal, although they do not use quasi-experimental methods to identify causal effects. In this context, we are the first to use Roman roads to construct instruments for measures of market potential in Portugal. While Roman roads have been used in more or less related studies for Spain (Garcia-López et al., 2015; Garcia-López, 2019; Holl, 2016), Italy (Percoco, 2016; De Benedictis et al., 2023; Bottasso et al., 2022), Germany (Wahl, 2017), and across Europe (Dalgaard et al., 2022; Flueckiger et al., 2022), no comparable application exists for Portugal. Second, we study firm location decisions from 1991 to 2016, during which road accessibility improved drastically, a much longer period compared to previous studies. Guimarães et al. (2000) and Holl (2004b, 2004c) investigated the determinants of firm location in the late 1980s to early 1990s, while Figueiredo et al. (2002), Melo et al. (2010) and Mota and Brandão (2013) focused on the 1990s and early 2000s.

Our results show that better road accessibility contributed to firm births, with estimated elasticities ranging from 1.6 to 1.9 for the 1-year interval and 1.2 to 1.3 for the 5-year interval. These results are robust to using an alternative set of IVs originated by enlarging the concept of non-locality and to changes in measures of the historical IVs. Although comparing results is not straightforward due to differences in model specification, these results are consistent with previous evidence for Portugal (Holl, 2004b, 2004c; Melo et al., 2010) and Spain (Holl & Mariotti, 2018b). Interestingly, the effects on the number of firm births are stronger when we exclude from the sample the two metropolitan areas of Lisbon and Porto, the most economically dynamic regions in mainland Portugal, suggesting a heterogeneous spatial effect. Interesting heterogeneous effects across economic sectors also emerge from the analysis. In particular, we find that the impact of market potential is primarily concentrated in the services sector, suggesting that firms from these industries, which rely heavily on proximity to final consumers, are highly responsive to changes in market potential and its impacts on firm birth. In contrast, no significant effects are observed in the manufacturing sector.

The remainder of the paper is organized as follows. Section 2 reviews the related literature. Section 3 discusses the empirical strategy and econometric modeling. Section 4 describes the dataset and provides descriptive statistics for the variables used in the empirical analysis. Section 5 reports and discusses the results, while Section 6 concludes.

## 2. Firm location, accessibility, and market potential

Several studies have tried to explain what drives firms' location decisions using either discrete choice models or count data models.[6] After Guimarães et al. (2003), count data models gained popularity due to the demonstrated equivalence between the likelihood function of discrete choice models (e.g., the Conditional Logit Model) and count data models (e.g., the Poisson Regression Model). The latter also offers the advantage of controlling for potential violations of the Independence of Irrelevant Alternatives assumption and reducing computational burden by treating the number of new firms choosing a specific location as a count variable rather than modeling a location choice indicator for each firm and location (Guimarães et al., 2004).

Among studies using count data models to investigate firm location, different explanatory factors have been considered. The literature distinguishes three approaches: neoclassical, institutional, and behavioral (Arauzo-Carod et al., 2010). Since our study is mainly interested in the role of improved road accessibility and market potential, we focus on typical neoclassical determinants. Neoclassical theories posit firms as rational agents optimally choosing locations to maximize profit or minimize costs. In this framework, some of the main determinants are, for instance, agglomeration economies, transport cost, and human capital.

Regarding transport, the literature suggests that better transport accessibility positively influences firm location, especially motorways (Alañón-Pardo & Arauzo-Carod, 2013; Holl, 2004a, 2004b, 2004c; Kim et al., 2018; Melo et al., 2010). Locating near transport infrastructure improves access to potential resources available to firms,

[6] For recent reviews, see Balbontin and Hensher (2019) and Carpenter et al. (2021).

affecting their costs and profits (Holl, 2004b). Nonetheless, a common problem these studies face is how they measure transport costs. With some exceptions (Gibbons et al., 2019; Gibbons et al., 2024; Holl, 2004a, 2004b, 2004c; Holl & Mariotti, 2018a), transport cost, or accessibility, is measured using straight-line distances or a dummy variable for the presence of transport infrastructure (Bhat et al., 2014; Daunfeldt et al., 2013; Kim et al., 2018). Therefore, they do not capture real accessibility using transport networks, nor, in particular, changes in real accessibility due to improvements made to transport networks.

Recent literature on Portugal examines the economic impact of introducing tolls on former toll-free motorways. During the 2010-2011 sovereign debt crisis, the Portuguese government implemented tolls in previously toll-free motorways to reduce public spending and increase revenue. Audretsch et al. (2020) find that tolls led to a 1.5% decline in the number of firms and a 3.5% drop in private sector nonfinancial employment in the municipalities affected by the introduction of tolls. In a related study, Branco et al. (2023) found that tolls decreased business turnover by 10.2% and full-time paid employment by 1.6% for firms located in municipalities where tolls were introduced. Although they sometimes explore the effects on firm births, these studies focus on different treatments and outcomes. An essential factor distinguishing our study from these is that we employ market potential to measure road accessibility – more on this below. This measure incorporates most of the road network, not just motorways. Additionally, we use accessibility data for 1991, 2001, and 2011, which mostly predates the shock introduced by tolls.[7]

As mentioned, a popular measure of transport accessibility is based on market potential, dating back to Harris (1954). The concept has more recently been discussed in the context of the so-called wider economic effects of transport improvements (Graham & Van Dender, 2011; Holl, 2012, 2016; Maré & Graham, 2013). Typically, this wider effect is measured by market potential or effective density indicators that include the compound effect of *economic size* (e.g., population, jobs) and *accessibility* (e.g., cost or time of traveling from origin to destination). Therefore, market potential reflects not only reduced travel times but also the benefits of improved access to larger markets, knowledge spillovers, and labor pooling.

[7] For further detail on market potential, please refer to Section 4.3.1.

Few studies estimate the impact of market potential on firms' location decisions. Holl (2004b) finds that improved market accessibility positively affects the spatial patterns of manufacturing firms in Portugal. In a related study, Holl (2004c) explores the effects of an increase in market potential across various sectors and finds mixed results. For instance, manufacturing sectors with higher transport costs benefit more, while those with lower transport costs prefer peripheral regions. Melo et al. (2010) also investigate how proximity to major economic centers affects firm location in Portugal. They find that doubling a municipality's market potential significantly increases the number of births, depending on the economic sector. Otsuka (2008) focuses on market accessibility and finds a positive effect on new firms across manufacturing and service sectors in Japan. More recently, Holl and Mariotti (2018b) show that improved access to transport infrastructure (road distance to the nearest motorway, airport, and seaport), market potential, and proximity to urban areas positively impact the location of logistics firms in Spain.

## 3. Empirical methodology

### 3.1. Count data model

We use the PPML estimator with multiple high-dimensional FE developed by Correia et al. (2020) to estimate models with the following specifications:

$$\begin{aligned}\mathbb{E}(n_{jst}) = \exp\big(&\beta_o + \beta_1 \log(MP_{j,t-1}) + \beta_2 \log(HHI_{j,t-1}) + \beta_3 LOC_{js,t-1} \\ &+ \beta_4 \log(WAGE_{js,t-1}) + \beta_5 \log(POP_{j,t-1}) \\ &+ \beta_6 \log(HC_{j,t-1}) + \alpha_j + \alpha_{s,t}\big) \qquad (1)\end{aligned}$$

$$t = \{1992, 2002, 2012\}$$

where $n_{jst}$ is the number of new firms in industry $s$, location $j$ and time $t$; $\mathbb{E}(n_{jst})$ stands for the conditional expectation of $n_{jst}$; $MP_{jt}$ is the market potential and $\beta_1$ is the coefficient of interest; $HHI_{jt}$ is the Herfindahl-Hirschman index; $LOC_{jst}$ is the specialization index; $WAGE_{jst}$ represents average real wage; $POP_{jt}$ represents population size; $HC_{jt}$ is the percentage of the population with complete higher education; $\alpha_j$ and $\alpha_{st}$ are municipality FE and FE for each combination of time and sector, respectively. The former, $\alpha_j$, controls for all unobserved location-specific characteristics and the latter ensures comparability between the Conditional Logit Model and the Poisson Regression Model as well as controlling for common factors specific to each combination of time and sector.[8]

The econometric literature shows that, under very general conditions, PPML is consistent, robust to overdispersion, and well-behaved even when the dependent variable shows a large proportion of zeros (Santos Silva & Tenreyro, 2006, 2011, 2022). Moreover, the PPML estimator is valid under a set of less restrictive assumptions than the Negative Binomial Regression Model. Another reason to use PPML regards panel data. On the one hand, as already discussed, although commonly used in firm location literature (Holl, 2004b, 2004c; Mota & Brandão, 2013), it is now well established that classic FE Negative Binomial Regression Model estimated by conditional Maximum Likelihood is not consistent, that is, "it does not necessarily remove the individual fixed effects in count panel unless a very specific set of assumptions are met" (Guimarães, 2008, p. 63). More recently, Blackburn (2015) reiterated this result. On the other hand, as noted by Santos Silva and Tenreyro (2022), since Wooldridge (1999), it is known that PPML does not suffer from the incidental parameter problem when considering one multiplicative unobserved effect with fixed T; Weidner and Zylkin (2021) have recently shown this property of PPML for gravity models with three-way FE. For these reasons, PPML estimator is now widely used in a broader range of economic applications (Santos Silva & Tenreyro, 2022), not only trade, but also, for example, tourism (e.g., Rosselló-Nadal & Santana-Gallego, 2024), migration flows between countries (e.g., Beine et al., 2016) and, as in our case, firm location decisions (e.g., Coll-Martínez et al., 2022; Moeller, 2018; Schlegel & Backes-Gellner, 2023).

### 3.2. Addressing potential endogeneity bias

Motorways or, more generally, transport infrastructure is not distributed randomly. For example, transport investment may be allocated to municipalities as a function of their present or expected socio-economic performance (e.g., higher population

8 In the context of panel data, the compatibility between the Conditional Logit Model and the Poisson Regression Model requires the inclusion of dummies for each combination of time and sector instead of dummies for each sector only (Guimarães et al., 2004).

growth or productivity rate), which is partially reflected, for instance, in the number of new firms. In most studies that use count data models to investigate firm location, endogeneity is addressed by including location-specific FE to proxy for omitted variables (Daunfeldt et al., 2013; Holl, 2004b, 2004c; Holl & Mariotti, 2018b). However, employing FE within the Negative Binomial Regression Model framework is generally discouraged (Guimarães, 2008), as discussed above. Additionally, some authors (Alañón-Pardo & Arauzo-Carod, 2013; Holl, 2004a; Melo et al., 2010) argue that the use of small geographical units reduces the likelihood of endogeneity bias by reverse causation because transport investment decisions are made at higher jurisdiction levels and thus should be exogenous to the local level. In our view, this argument may not be entirely valid. Despite political decisions regarding the allocation of new investments being made at central or regional levels of government, the specific characteristics of municipalities can still influence the placement of transport investment. For instance, political economy factors, such as the alignment between local and national governments, can influence the location of investment at the municipal level (Rocha et al., 2022).

The limited discussion of endogeneity bias in the firm location literature is somewhat surprising. It contrasts with the broader literature on the relationship between the spatial organization of economic activity, transport infrastructure, and agglomeration economies (for a review, see Redding & Turner, 2015). In this broader literature, using historical and planned transportation networks to instrument endogenous modern transportation networks is now standard. Baum-Snow (2007) was the first to implement an IV strategy based on planned routes. The author used the number of motorways in the 1947 National Interstate Motorways Plan as an IV for the number of motorways built between 1950 and 1990. Duranton and Turner (2011, 2012) used, among others, the routes of the major expeditions of the U.S. between 1535 and 1850 as sources of quasi-random variation for the U.S. interstate motorway network at the end of the 20th century. A series of studies by various authors followed, employing similar types of IVs. Agrawal et al. (2017) and Heidt and Kasim (2020) utilized historical maps and plans to instrument for modern infrastructure in the U.S., while Baum-Snow et al. (2017) applied mid-20th-century IVs for China. More recently, Rocha et al. (2023b) focused on 18th-century dirt roads and the 1945 National Road Plan for Portugal.

One possible explanation for the absence of the identification strategy mentioned above might be related to the non-linear nature of count data models. While the IV method is straightforwardly applied to linear models to correct for endogeneity, its application to non-linear models is more complex. To our knowledge, Alañón-Pardo and Arauzo-Carod (2013) is the only study based on count data models that attempt to use IV. However, they employ a relatively unusual approach, substituting the endogenous variable with a proxy variable assumed to be exogenous. The authors use population change between 1970 and 1991 as their proxy variable, or instrument for motorways in the 1990s, and include it directly in their Negative Binomial Regression Model.

This study is – to the best of our knowledge – the first to employ what is now standard in the broader literature mentioned above, namely the use of historical information to construct our IV, in combination with the PPML technique (Santos Silva & Tenreyro, 2006, 2011, 2022). Specifically, we attempt to estimate the causal effect of road accessibility, through market potential, on firm births using a PPML estimator, combining IV and FE methods with the control function approach proposed by Lin and Wooldridge (2019).[9] This technique employs a two-stage procedure to estimate an IV in a non-linear model context. The first stage, estimated using Ordinary Least Squares (OLS), uses the endogenous variable as the dependent variable and includes the IV, control variables, and FE. The second stage, estimated with PPML, includes the residuals from the first stage along with the endogenous variable, control variables, and FE. By including the residuals from the first stage, we expect to control for the variation in market potential (suspected to be endogenous) not explained by the IVs. As a result, the estimated coefficient in the second stage measures the effect of market potential on the count of plant birth after netting out endogeneity. For a more detailed and technical explanation, see Appendix B.

#### 3.2.1. Identification strategy

We aim to minimize bias due to the endogeneity of market potential through a comprehensive strategy involving IV, FE, and control variables, in addition to using lagged values of the explanatory variables. Including municipality FE allows us to control for any correlation between time-variant explanatory variables and time-constant

[9] This technique has recently been applied in other studies, e.g., Hidalgo et al. (2024), Miroudot and Rigo (2022), and Nocito et al. (2023).

unobserved heterogeneity. Yet, in many contexts, the presence of a correlation between covariates and the error term can introduce endogeneity bias that FE alone cannot address. For those cases, one possibility is to use IV. Hence, combining IV and FE gives us additional confidence that the estimated coefficients may, indeed, reflect causal effects. In addition to municipality-level FE and FE for each combination of time and economic sector, we also expect to attenuate omitted variables bias by controlling for the five location factors described in Section 4.3.2. In this respect, given that our approach for addressing endogeneity involves IV, the choice of control variables is crucial to ensure the exogeneity of the excluded IVs (Duranton & Turner, 2012) – more on this below.

Additionally, to further attenuate simultaneity bias, we compute the dependent variable as the total of the five years after the year the explanatory variables were observed (hereafter, we refer to this as the 5-year interval to differentiate from the 1-year interval considered previously). Specifically, the panel data based on 1-year intervals comprises the number of new establishments in 1992, 2002, and 2012. For the 5-year interval, we consider the sum of new establishments for 1992-1996, 2002-2006, and 2012-2016. In both cases, explanatory variables are measured in 1991, 2001, and 2011. By examining longer periods, we expect to avoid year-specific abnormalities (e.g., unusually high or low numbers of births in a given year) and also strengthen the argument for non-reverse causality.

Note that when computing market potential, we use the state of the road network as observed in 1991, 2001, and 2011. This means that many reductions in travel times resulting from improvements in the network (e.g., the expansion of the motorway network) occurred before these specific years, which also contributes to eliminating simultaneity bias. For example, there might have been network expansions in 2000-2001 or well before that, e.g., 1992-1993 – both will be captured by the market potential variable in 2001, which, in our models, will explain firm births in 2002 or 2002-2006.

Finally, it is worth making two observations regarding our empirical methodology. First, by employing the PPML estimator, we depart from the context of the well-known two-stage least squares (TSLS) estimator. Instead, we implement the control function approach introduced by Lin and Wooldridge (2019) to estimate Equation (1) – hereafter, PPML-FEIV. Second, employing FE and IV simultaneously necessitates time-variant instruments. We elaborate on this below.

### 3.2.2. Composite time-variant instruments

Since, by definition, the IVs based on historical roads are not time-variant, we cannot use them in combination with FE. To overcome this limitation, we follow a similar approach to Garcia-López et al. (2015), Garcia-López (2019), Holl (2012), and Holl and Mariotti (2018a) and interact historical roads with a modern non-local road network measure that varies in time. Essentially, this approach involves constructing a CIV composed of two elements: a time-variant component expressed by a measure of the modern road network and a time-invariant component expressed by historical information. Our approach is also inspired by Ignatov (2023), who exploits non-local motorway improvements as a source of exogenous variation for market access.

Among these studies, Garcia-López (2019) referred to their CIV as a "time-variant historical instrument". Here, as our (endogenous) variable of interest, the market potential is essentially a non-local variable, we propose a "non-local time-variant historical instrument". More specifically, for our two historical components (described in detail in Section 4.4), we first consider the Roman major roads, based on the data from McCormick et al. (2013), which have been widely used in the literature to instrument modern transport infrastructure (Dalgaard et al., 2022; De Benedictis et al., 2023; Flueckiger et al., 2022; Garcia-López, 2019; Wahl, 2017). For our second historical component, we follow Rocha et al. (2022, 2023, 2024) and use the maps of the 18th-century itineraries.[10] As discussed in Section 3.2, maps from this period (18th to 19th centuries) are also widely used in the literature as IV for modern transport infrastructure. For example, Duranton and Turner (2012) employed maps of early explorations (1528-1850) and an 1898 railroad map as IVs for motorways in the US. Similarly, Garcia-López et al. (2015) and Garcia-López (2019) used the 1760 Bourbon roads and the 1870 railroad network as IVs for motorways in Spain, respectively. For Germany, Möller and Zierer (2018) relied on historical IVs using an 1890 railroad plan, while Levkovich et al. (2020) used 1821 road maps as IVs for motorways in the Netherlands. For both historical components, we consider the

[10] As our primary focus is on the Roman roads of the Iberian Peninsula, particularly Portugal, we also used data from a new broader project called *Mercator-e*, which centers on the Iberian Peninsula and its recent extension to cover the entire Roman Empire, known as *Itiner-e*, to test the robustness of our results. Unlike the McCormick et al. (2013) dataset, the *Mercator-e* and *Itiner-e* data have not yet been widely used in the literature. These projects have developed extensive raw data related to historical routes, representing what is likely the most comprehensive compilation of Roman routes to date, especially for the Iberian Peninsula (Brughmans et al., 2024; De Soto & Carreras, 2021; De Soto, 2019). Maps and results using this recent information are reported in Appendix E.

inverse of the mean straight-line distance to the nearest Roman major roads and the 18th-century itineraries, computed from the municipalities within a non-local region discussed in more detail below.

Concerning the time-variant element of our CIVs, we use the NUTS2 regions to construct a non-local area to calculate the ratio of the length of motorways and expressways over the national network in that year. Since NUTS2 regions might include areas with characteristics similar to those of the municipalities and thus be influenced by potential sources of endogeneity, we exclude not only the municipality's own contribution but also that of its neighbors from the computation of our CIVs. This defines the non-local area as the remaining municipalities in that NUTS2 region, which we use to compute both components of our CIV.[11] Formally, our CIV can be described as follows:

$$CIV_{j,t} = \frac{1}{\bar{d}_j} \times \frac{\sum_{l \in R(j) \setminus G(j)} mw_{l,t}}{\sum_{i \in N \setminus G(j)} mw_{i,t}} \tag{2}$$

$$\bar{d}_j = \frac{\sum_{l \in R(j) \setminus G(j)} d_l}{card(R(j) \setminus G(j))}$$
$$G(j) \subseteq R(j) \tag{3}$$
$$R(j) \subseteq N$$

where $d$ is the straight-line distance of the municipality centroid to the nearest historical road (i.e., the Roman major roads or the 18th-century itineraries), in kilometers; $\bar{d}$ refers to the mean distance; $mw$ refers to the length of motorways and expressways, in kilometers; $G(j)$ refers to the set of municipality $j$ and its neighbors; $R(j)$ refers to the set of municipalities located in the same NUTS2 region of municipality $j$; $N$ refers to the set of all the 275 municipalities in mainland Portugal; and $card$ refers to the cardinality (or the number of elements) of a set. Since the distance to the nearest historic IV is expected to affect market potential negatively, we use the inverse distance for a more direct interpretation in the first stage. As can be seen in the equation above, non-locality is reflected in both components of our CIVs, as historical information is aggregated within

[11] Note that we also exclude the contributions of the municipality and its neighbors when calculating the length of the national network, which is the denominator of the time-variant element of our CIV – see the second term of Equation (2).

each NUTS2 region, excluding the municipality's own contribution and that of its neighbors.

The validity of these IVs requires that they must be relevant, i.e., they should be a good predictor of the municipality's market potential.[12] In this regard, we expect the time-variant component of our CIV to be correlated with market potential, mainly through the accessibility component. Road transport improvements affect market potential by reducing costs associated with distance between locations, facilitating firm access to different markets, thereby increasing market potential. Additionally, similar to market potential, the ratio of the length of motorways and expressways in the non-local area to the national network serves as a measure of (non-local) road accessibility. The non-local area must be sufficiently distant from a municipality $j$ to enhance the likelihood of exogeneity, yet not so far that the instrument becomes only marginally related to the market potential of $j$, which would result in a weak instrument. An increased road ratio in non-local areas implies a higher degree of connectivity and access for municipalities within these regions, thereby enhancing firms' capacity to reach diverse markets. This component is expected to be directly related to the accessibility aspect of market potential.

Regarding historical IVs, it is based on the idea that there is a positive correlation between the modern road network and the historical network due to common factors that have influenced their construction, which may remain relatively constant over time – hence the necessity to control for municipality FE in our regressions. This type of path-dependence may reflect not only construction costs and feasibility (for example, due to local topographical features, such as terrain ruggedness, presence of rivers, etc.) but also historically persistent socioeconomic factors (e.g., settlement patterns around defensive structures). Factors of this type may have driven transport infrastructure development in similar locations at different historical periods (Rocha et al., 2022). Thus, it is reasonable to expect that proximity to these historical routes significantly increased the likelihood that a modern motorway would later be built in the same area, thereby enhancing road accessibility and, consequently, market potential.

It is also required that the IVs are not correlated with the error term, conditional on controls. That is, the exclusion restriction requires that CIV can only affect firm births through market potential, conditional on controls. As for the historical elements of our

[12] Refer to the lower part of Table 3 in Section 5.1 or similar tables in Section 5.2 for the first-stage statistics.

CIV, it does not seem very likely that, conditional on controls, the Roman major roads and the 18th-century itineraries could be systematically correlated with the current spatial distribution of firm births. The Roman roads and the 18th-century itineraries are, respectively, more than a thousand years and more than two hundred years apart from 1991, the initial year in our analysis.[13] Over the past decade, numerous studies have employed Roman roads as an instrument for modern motorways, highlighting its validity in terms of both relevance and satisfying the exclusion restriction (see, for instance, Bottasso et al., 2022; Holl, 2016; Percoco, 2016; Roca & Puga, 2017). Regarding the latter, the literature identifies the strategic role of Roman roads in facilitating military movements as the primary motivation for their construction, particularly for the major roads (De Benedictis et al., 2023; Garcia-López et al., 2015; Santagata, 2022). It has also been noted that even after their construction, the economic impact of Roman roads was limited, as rivers and sea routes remained the most efficient and cost-effective means of transporting goods (Finley, 1973, as in De Benedictis et al., 2023).

Similarly, the 18th-century itineraries in Portugal were not considered efficient transportation routes. Inland waterways and coastal navigation were viewed at the time as the country's primary transport networks and the best options for long-distance goods transport (Rocha et al., 2022, 2023, 2024). Dirt roads primarily served short-distance travel and were often in poor condition, especially during winter when rain often made many routes impassable due to mud (Link, 1803; Matos, 1980; Justino, 1988, as cited in Rocha et al., 2024). These routes were primarily used for travel by foot, horseback, or donkey, and in 1810, it would take an express courier around three days to journey from Lisbon to Porto – approximately 313 km (Matos, 1980, as cited in Rocha et al., 2024). As a result, travel by these roads was often impractical.

These temporal gaps and the historical context of road construction (such as military purposes and the constraints imposed by poor conditions and limited usability) suggest that these roads were not built in ways that anticipated contemporary factors directly affecting firm location decisions. Furthermore, as noted earlier, in addition to conditioning on controls, geographical factors or any other historical characteristics – whether observed or unobserved – that might influence firm births through channels other than market potential are controlled for by including FE.

[13] In fact, these itineraries reflect essentially roads in 1748 (Rocha et al., 2024).

Regarding the time-variant component of our CIVs, we argue that by using the large NUTS2 regions and excluding not only the municipality's own contribution but also that of its neighbors from the computation, we expect to provide exogenous variation of road accessibility. In mainland Portugal, municipalities are relatively large compared to several other European countries, covering an average area of about 324 km², significantly larger than, for instance, France (15 km²), Germany (30 km²), Italy (37 km²), and Spain (62 km²). Since each municipality has five neighbors on average, this results in excluding an area of approximately 1,944 km² – about 10% of the average area of the NUTS2 regions. By excluding a local area of this size, we aim to base our time-variant component on road improvements in relatively distant (non-local) areas, thereby providing, in principle, exogenous variation to study the impact of better road accessibility on firm births.

Given that most EU funds were allocated to municipalities within NUTS2 regions, we use the 1986 NUTS2 classification to rule out the possibility of endogeneity from using the current NUTS2 classification, in place since 2013.[14] After joining the EU, Portugal adopted the NUTS classification, which underwent successive changes. Many of these changes rearranged municipalities from the richer NUTS2 region of Lisbon and Vale do Tejo to the poorer NUTS2 regions of Alentejo and Centro, precisely to increase the influx of EU funds to these regions. Since a significant share of the road investment was based on European structural funds, the current NUTS2 classification will likely be affected by endogeneity, justifying our use of the 1986 classification. There were five NUTS2 regions in mainland Portugal in 1986, each with an average of 55 municipalities. For a comparison between classifications, see Figure C1 in Appendix C.

Finally, as previously discussed, ensuring the exogeneity of excluded IVs relies on including the appropriate set of controls. It would be reasonable to assume that the control variables described in Section 4.3.2 could influence both IVs and the dependent variable (e.g., economic performance in a region can be affected by sectoral specialization, thereby influencing policymakers to improve transport infrastructure to boost growth or the number of new firms, which can take advantage of productivity gains) or represent other channels through which excluded IVs may affect the dependent variable (e.g., transport investment may influence the number of firms through population

[14] After completing this manuscript, a new NUTS2 division was established in 2024, expanding the previous five-unit classification in 2013 to seven units in the new classification.

size, as this type of investment could be assigned to a region to anticipate the needs of a growing urban population, thereby influencing the number of firms through clustering in this area). Including these control variables in our empirical model enhances the reliability of the exogeneity of the IVs and, consequently, strengthens the identification of a causal effect.

## 4. Data and descriptive statistics

### 4.1. Firm- and establishment-level data: Defining births.

Our main data source is the survey *Quadros de Pessoal*, produced by the Ministry of Labor and Social Solidarity (*Ministério do Trabalho, Solidariedade e Segurança Social*), covering all private sector firms in mainland Portugal and the Autonomous Regions of Azores and Madeira with at least one paid employee. Conducted annually since 1981, data is available for research from 1986 onwards.[15] Since 2010, only employer firms have been required to respond, changing the population covered compared to previous years when both employer and non-employer firms had to respond. The survey provides detailed information on establishment location, sector of economic activity, and workforce characteristics such as education level and wage. Firms, establishments, and employees each have unique identifier codes. These codes enable the mapping of all new establishments.

The literature defines firm births as those appearing for the first time in a given region and year. We focus on employer firms, i.e., enterprises with at least one paid employee, as done, for example, by Anyadike-Danes and Hart (2018), Coad et al. (2018), or the Eurostat and OECD, which use the concept of *employer enterprise birth* (Eurostat-OECD, 2007). For consistency, we define births as establishments appearing for the first time in the database with at least one paid employee in a given year. We analyze the firms located in mainland Portugal only because the main interest is the effect of better road accessibility. In contrast, air and sea infrastructure are likely to play, in relative terms, a more relevant role in the Islands. Several data cleaning and harmonizing procedures were carried out to combine each cross-section into a longitudinal dataset. These included

[15] In 2001, detailed workforce information was unavailable, affecting wage data only. To address this missing data issue, the average wage for 2001 is computed by taking the mean value of wages from 2000 and 2002.

eliminating observations referring to two-digit economic sectors that were not clearly defined and which we could not match across periods using the classifications of economic activity in place between 1992 and 2016. See Appendix A for the final list of the 32 economic sectors and more information on the harmonization procedure.

After data cleaning, the final dataset includes information on 107,775 new establishments in mainland Portugal for 1992, 2002, and 2012, and 520,708 new establishments recorded over the 5-year periods of 1992-1996, 2002-2006, and 2012-2016. We aggregate this establishment-level data at the municipality and sector levels to construct our dependent variable, the total number of new firms by municipality and sector, resulting in a panel dataset with a total of 26,400 observations.

### 4.2. Firm births: Spatial analysis and descriptive statistics

Table 1 shows descriptive statistics for the count of births per municipality for 1992, 2002, and 2012 (Part A) and the total number of births for 1992-1996, 2002-2006, and 2012-2016 (Part B). In 1992, Lisbon recorded the highest number of 4,121 births, while between 1992 and 1996, it reported a total of 17,626. Lisbon consistently experienced the highest values throughout the years analyzed, both for the single years 1992, 2002, and 2012, and over the 5-year periods. Virtually no municipality recorded zero births, with two exceptions: Valpaços and Alvito in 1992 and 2012, respectively.

**Table 1. Descriptive statistics: Count of new firms in Portuguese municipalities**

| A. | 1992 | 2002 | 2012 |
|---|---|---|---|
| | Count of firm births | | |
| Min | 0 | 4 | 0 |
| Mean | 127.01 | 177.07 | 87.83 |
| S.D. | 301.38 | 324.13 | 185.69 |
| Max | 4,121 | 3,879 | 2,368 |
| Total | 34,928 | 48,693 | 24,154 |

| B. | 1992-1996 | 2002-2006 | 2012-2016 |
|---|---|---|---|
| | Count of firm births | | |
| Min | 13 | 43 | 12 |
| Mean | 608.00 | 782.49 | 503.00 |
| S.D. | 1,334.27 | 1,402.45 | 1,086.24 |
| Max | 17,626 | 16,908 | 14,216 |
| Total | 167,199 | 215,184 | 138,325 |

**Notes.** Part A shows the count of births per municipality for the years 1992, 2002, and 2012; Part B shows the total number of births for the periods 1992-1996, 2002-2006, and 2012-2016; "S.D." stands for standard deviation; "Min" and "Max" stands for minimum and maximum, respectively.

**Source:** Authors' calculations.

Figure 1 illustrates the spatial distribution of new firms in 1992, 2002, and 2012**Table *1***. Births are concentrated on the west coast of mainland Portugal, especially within the Lisbon metropolitan area (LMA) and Porto metropolitan area (PMA), but also in some district capitals in the interior of the country.[16] This picture is not surprising since the two metropolitan areas are the country's most populated and economically important regions. The LMA and PMA comprise only 17 and 16 municipalities out of 275, respectively.[17] Nevertheless, in 2021, they represented 46.7% of the Portuguese (mainland) population, 54.6% of jobs in non-financial enterprises, and 54.1% of GDP.[18]

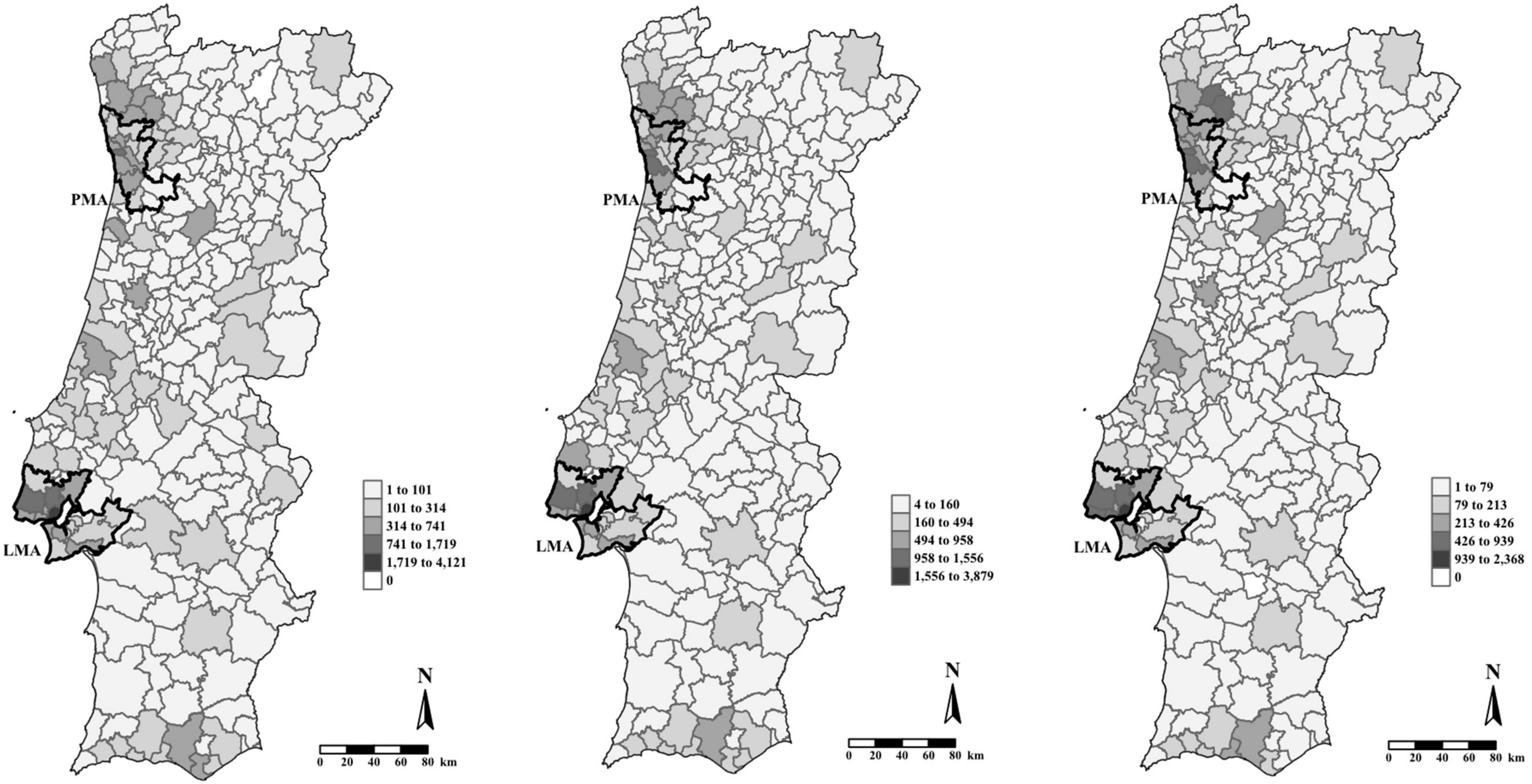
---

[16] The LMA and PMA are the only two metropolitan areas in Portugal.

[17] Mainland Portugal has 278 municipalities since 1998, but we use the pre-1998 administrative division of 275 municipalities to ensure data consistency. Consequently, the LMA and PMA consist of 17 and 16 municipalities rather than 18 and 17, respectively.

[18] See PORDATA/pop; PORDATA/nonfinancial_jobs; and PORDATA/gdp.

**Figure 1. Spatial distribution of firm births by municipalities**




**Panel A. Number of births in 1992.** **Panel B. Number of births in 2002.** **Panel C. Number of births in 2012.**

**Notes**. "PMA" stands for Porto Metropolitan Area; "LMA" stands for Lisbon Metropolitan Area; the LMA and PMA are the only two metropolitan areas in Portugal. Maps were generated based on the "Jenks classification" in R, which identifies logical breakpoints by grouping similar values that minimize differences between values in the same cluster and maximize the differences between clusters. **Source:** Authors' calculations.

### 4.3. The determinants of firm location

Building on previous studies, we summarize the most common determinants of firm location in three groups: transport accessibility, agglomeration economies, and labor market characteristics. Appendix Table C1 describes the variables used to capture the effect of different location determinants, descriptive statistics, and previous count data studies that have used these variables or similar measures. Below, we provide detailed descriptions of market potential, our primary variable of interest, and other variables included as controls.[19]

#### 4.3.1. Market potential and road accessibility

As noted in Section 2, market potential includes the compound effect of economic size and the cost incurred to access it, usually measured using distances or travel times between origin-destination pairs (Graham, 2007). Formally, market potential is calculated as:

$$MP_j = \sum_{k=1,k\neq j}^{275} \frac{POP_k}{d_{jk}} \quad (4)$$

where $POP_k$ is the population of the destination $k$. Using a geographic information system (GIS), $d_{jk}$ is calculated as the travel time by road between the population-weighted centroids of municipalities $j$ and $k$ (Afonso et al., 2025).[20] The measure of market potential captures all improvements in the road network. Given prior research (Holl, 2004a, 2004b, 2004c; Holl & Mariotti, 2018b; Melo et al., 2010; Otsuka, 2008), we expect market potential to impact firm location decisions positively.

The development of modern roads in Portugal did not begin until the mid-20th century.[21] As Pacheco (2004) noted, it was only after the country's accession to the EU in 1986 that the road network experienced significant modernization and expansion to

[19] The market potential and other control variables are constructed or measured for the census years 1991, 2001, and 2011.
[20] We adopt the standard value of 1 for the spatial distance decay parameter, commonly used in the literature. Evidence suggests that alternative values do not substantially alter the general conclusions. For example, Gibbons et al. (2019) explore various distance decay values and report that these alternatives yield highly correlated accessibility indexes with no significant changes to the main findings.
[21] See Rocha et al. (2022) for a primer on the recent road developments in Portugal.

improve inter- and intra-regional accessibility. Despite initially lagging behind other European countries until the mid-1980s, the expansion of the Portuguese motorway network was remarkably fast (Rocha et al., 2022). As mentioned, the motorway network increased from 211 km in 1986 to 3,065 km in 2014. Figure 2 shows the expansion of the road network, particularly motorways and expressways, for the years 1991, 2001, and 2011. This substantial improvement of the road network is expected to have a predominant impact on market potential. Table 2 presents market potential, population, and road network summary statistics. For instance, the average bilateral travel time (the travel time distance between any two municipalities) decreased from approximately 192 minutes in 1991 to 151 minutes in 2011, reflecting a reduction of 21.4%. To specifically quantify how much of the variation of market potential is attributed to changes in the road network (which we term the *road effect*) or to population changes (the *population effect*), we decompose the variation in market potential using an approach à la shift-share for each municipality $j$, as presented below:[22]

$$\frac{\Delta MP}{MP^{t-1}} = \frac{MP^{t} - MP^{t-1}}{MP^{t-1}} = \frac{\Delta \overline{MP}}{MP^{t-1}} + \frac{\Delta \widetilde{MP}}{MP^{t-1}} + \frac{\Delta \epsilon}{MP^{t-1}} \quad (5)$$

where $MP$ represents market potential calculated in Equation (4); $MP^{t-1}$ represents the initial value of $MP$ (i.e., values for 1991 or 2001); $\overline{MP}$ represents market potential calculated using population of 1991 fixed – in this case, the variation is solely attributed to changes in the road network; $\widetilde{MP}$ represents market potential calculated using the road network of 1991 fixed – in this case, the variation is solely attributed to changes in population; and $\epsilon$ is a residual term.[23] Each term of Equation (5) was calculated individually and then averaged across municipalities. Table 2 shows that $\overline{MP}$ accounts for 74.7% and 77.7% of the variation in 1991-2001 and 2001-2011 period, respectively, while $\widetilde{MP}$ accounts for only 21.2% and 13.7% in the same periods.[24] This suggests that increases in market potential are primarily due to increases in road accessibility.

[22] For a proof of the decomposition, please refer to Appendix B.

[23] Our analysis uses theoretical travel times, which do not account for traffic congestion or other impedances. These travel times represent optimal conditions under free-flow scenarios. As a result, the road accessibility component of the market potential measure is not influenced by congestion effects induced by large population sizes. Using theoretical travel times ensures that the distinction between road and population effects remains valid.

[24] Specifically, from (3) in Table 2, we divided 18.05 (second row) by 24.15 (first row) in column 2 and 7.68 (second row) by 9.89 (first row) in column 3, resulting in percentages of 74.74% and 77.67%, respectively.

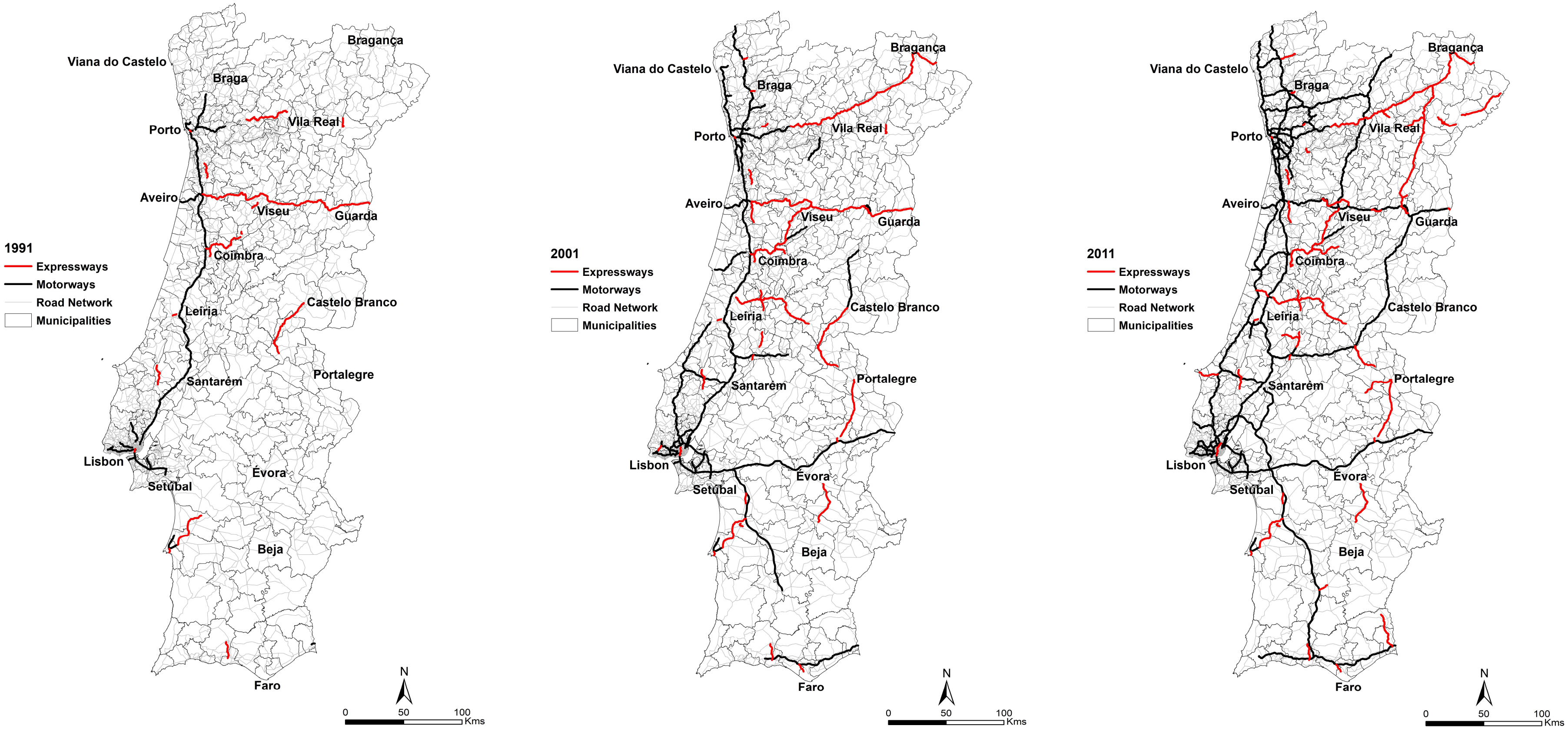
**Table 2. Summary statistics of market potential and road accessibility**

| (1) Summary statistics | 1991 | 2001 | 2011 |
|---|---|---|---|
| Total length of motorways and expressways (km) | 994 | 2,583 | 3,985 |
| AAGR (%) | n.a. | 10.02 | 4.43 |
| Average bilateral travel time (min) | 192 | 164 | 151 |
| AAGR (%) | n.a. | -1.42 | -0.75 |
| Average population across municipalities | 34,094 | 35,889 | 36,537 |
| AAGR (%) | n.a. | 0.47 | 0.16 |
| **(2) Average market potential** | | | |
| $MP$ | 79,079.64 | 97,200.98 | 106,460.77 |
| $\overline{MP}$ | 79,079.64 | 92,413.32 | 99,457.01 |
| $\widetilde{MP}$ | 79,079.64 | 83,145.03 | 84,540.35 |
| **(3) Average variation of market potential (%)** | | | |
| $\Delta MP/MP^{t-1}$ | n.a. | 24.15 | 9.89 |
| $\Delta\overline{MP}/MP^{t-1}$ | n.a. | 18.05 | 7.68 |
| $\Delta\,\widetilde{MP}/MP^{t-1}$ | n.a. | 5.13 | 1.36 |
| **(4) Share of the average variation of market potential (%)** | | | |
| $\overline{MP}$ | n.a. | 74.74 | 77.67 |
| $\widetilde{MP}$ | n.a. | 21.23 | 13.71 |
| $\epsilon$ | n.a. | 4.03 | 8.62 |
| | n.a. | 100 | 100 |

**Notes.** AAGR stands for "Average Annual Growth Rate"; $MP$ stands for market potential calculated as in Equation (4); $\overline{MP}$ stands for market potential calculated using the population of 1991 fixed; $\widetilde{MP}$ stands for market potential calculated using the road network of 1991 fixed; The values presented in (4) are obtained by calculating the ratio of the other two (second or third) rows to the first row in (3). **Source:** Authors' calculation.

Hereafter, we focus on the market potential calculated using the population of 1991 fixed, i.e., the *road effect*, as the primary measure of analysis, rather than the complete market potential measure – Equation (4), which is discussed in the robustness analysis section (see Section 5.2.2). As previously demonstrated, the shift-share approach enables us to isolate the component of market potential linked to improved road accessibility from that associated with population size (which, in principle, could constitute a source of endogeneity). Note, however, that the two measures are highly correlated, as shown in the correlation matrix (Table C2) and scatterplots (Figures B1 and B2).

## Figure 2. Motorways and expressways in mainland Portugal




**Panel A. Motorways and expressways in 1991.** **Panel B. Motorways and expressways in 2001.** **Panel C. Motorways and expressways in 2011.**

**Source:** Maps adapted from Afonso et al. (2025).

### 4.3.2. Control variables

*Agglomeration economies: Urbanization and localization*

Agglomeration economies are arguably the most common factor considered by existing studies of firm birth, and, in general, the results indicate that they play an important positive role (Bhat et al., 2014; Capozza et al., 2018; Guimarães et al., 2004; Jofre-Monseny et al., 2014). The literature usually distinguishes between urbanization and localization economies. According to Graham and Gibbons (2019), the former can be defined as externalities common to all firms, i.e., external to the firm and the industry but internal to the city, while the latter are externalities that benefit firms in the same industry, i.e., external to the firm but internal to the industry.

We use the Herfindahl-Hirschman index (HHI) to proxy for urbanization economies, varying between 0 and 1. A higher value indicates a greater concentration of economic activity in fewer sectors, implying lower diversity. If diversity positively influences firm location decisions, its coefficient is expected to be negative (Holl, 2004c). The index is computed as follows:

$$\mathrm{HHI}_j = \sum_s \left(\frac{e_{sj}}{e_j}\right)^2 \tag{6}$$

where $e_{sj}$ is the employment in sector $s$ in municipality $j$ and $e_j$ is the total employment in municipality $j$.

Localization economies are measured by an index of relative specialization (LOC), which captures the presence of firms in the same sector through local-national comparison. A value greater than 1 suggests a higher degree of relative specialization in a given sector. The specialization index is calculated as follows:

$$LOC_{sj} = \frac{e_{sj}/e_j}{e_{sn}/e_n}, \tag{7}$$

where $e_{sj}$ is the employment in sector $s$ in municipality $j$; $e_j$ is the total employment in municipality $j$, $e_{sn}$ is the national employment in sector $s$, and $e_n$ is the total national employment. We cannot anticipate a clear effect of own-sector specialization as results might be mixed (Holl, 2004c; Melo et al., 2010).

*Labor market factors: wages, human capital, and market size*

Since wages and workers' skills also affect a firm's production and costs, we use the average real wage for each municipality and sector of economic activity and the percentage of the total resident population with complete higher education as measures of labor cost and labor quality (i.e., human capital), respectively.[25] Average wages are included in almost every firm location study (Guimarães et al., 2000, 2004; Melo et al., 2010). Areas with higher average wages typically imply greater costs for firms, leading to a negative effect. The quality of the labor force is another major factor considered in several studies (Mota & Brandão, 2013). Municipalities with a more educated labor force are expected to be more attractive to firms, suggesting a positive effect. Another determinant considered in the literature is market size. We use the total resident population to measure local market size and expect a positive effect on firm location.[26]

### 4.4. Historical road network: The Roman major roads and the itineraries from circa 1800

The Romans were the first to build an extensive, sophisticated network of durable paved roads designed to last indefinitely (Percoco, 2016). The raw data on the Roman road network was digitized by McCormick et al. (2013) as part of the Digital Atlas of Roman and Medieval Civilization (DARMC) project by Harvard University's Center for Geographic Analysis.[27] This network passes through 36 countries across Europe, Africa, and Asia, with road segments primarily classified by their importance (e.g., major and minor roads). In total, 192,861 km of Roman roads are documented, of which 92,749 km

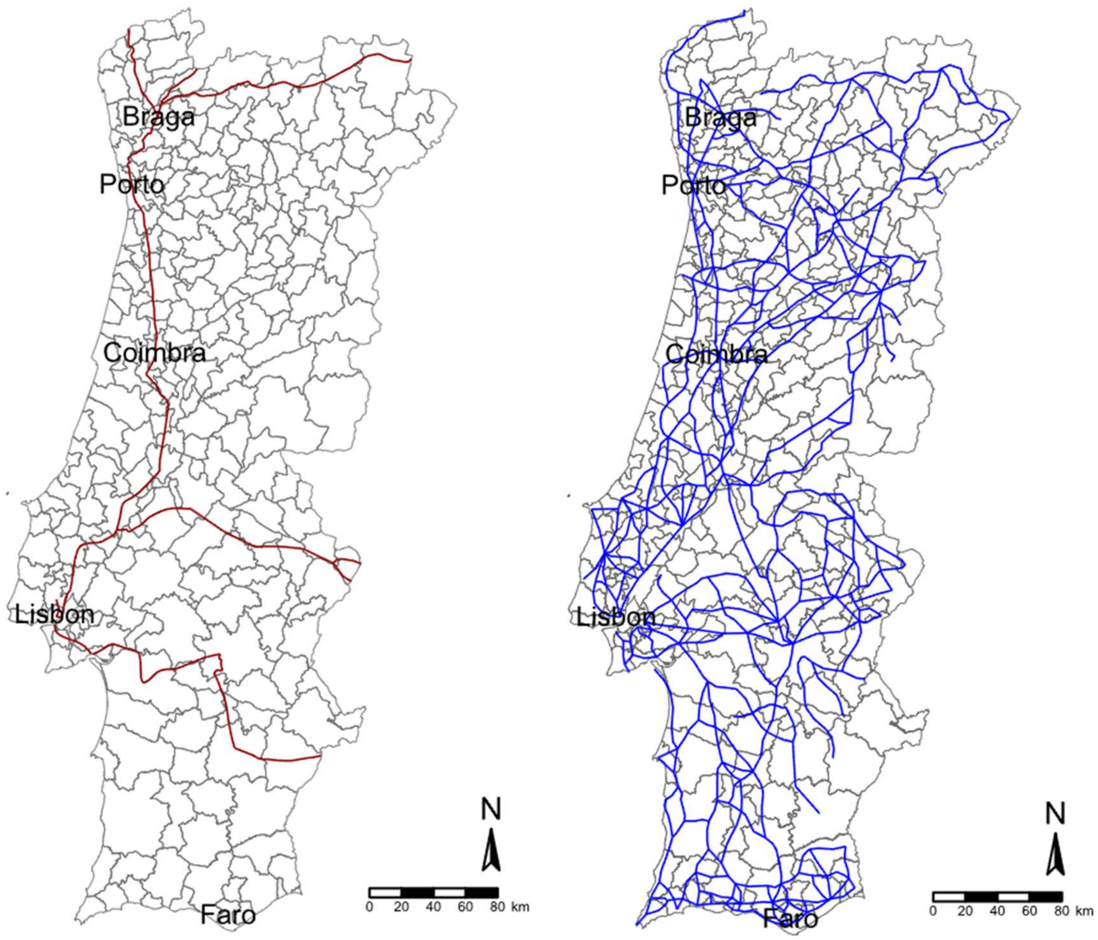
[25] In cases where no firms of a specific economic sector exist within a municipality, we compute the wage as the overall average for that municipality, regardless of sector.
[26] Some studies use population density as a proxy for land cost (e.g., Figueiredo et al., 2002; Guimarães et al., 2004; Mota & Brandão, 2013) and not as a proxy for market size. In that case, if firms compete for the same space, it is expected to affect firm location choice negatively.
[27] See DARMC.

are classified as major routes, essentially built for military reasons. For the purposes of this paper, we focus on the Portuguese segment of this infrastructure, where around 1,085 km of these major roads cross present-day Portugal.

The mainly unpaved "itineraries" of the 18th century formed a web of precarious roads that developed over the centuries without the centralized planning of a transport network and preceded by more than half a century the period in which roads began to be built regularly and systematically (Rocha et al., 2023). The primary source of data for these itineraries comes from Matos (1980). Although roads in this period were generally in poor condition and had limited use (as discussed in Section 3.2.2), they formed a relatively extensive and dense network. A total of 7,662 km of these routes existed, with the average distance from a municipality centroid to the nearest 1800 itinerary being no more than 3.2 km. Figure 3 illustrates both the Roman major roads (Panel A) and the 18th-century itineraries (Panel B). It also shows how these two historical networks differ in length, density, and distribution across mainland Portugal.[28]

[28] The cross-sectional correlation between these two historical road networks is relatively low, around 0.22.

**Figure 3. Historical roads**




**Panel A. The Roman major roads** **Panel B. The 18th-century itineraries**

**Source**: Authors' calculations based on data from McCormick et al. (2013) (Panel A) and from Rocha et al. (2022, 2023, 2024) (Panel B).

## 5. Results and discussion

### 5.1. Baseline results

Table 3 reports results from PPML-FE (Panel A) and PPML-FEIV (Panel B) estimations. At first sight, when considering the results from Panel A, we would conclude that the coefficient of market potential estimated ignoring endogeneity is not significant for the 1-year interval although having the expected positive sign, while for the 5-year interval, it is statistically significant at 1% with an elasticity of 0.59 (Column 2). However, given that market potential is most likely to suffer from endogeneity (as discussed in Section 3.2), estimates may suffer from a negative bias, explaining the lack of significance. Accordingly, we instrument market potential individually with the two CIVs (i.e., each model is exactly identified) described in Section 3.2.2: the ratio, for each year, of the length of motorways and expressways computed for the non-local area over the national level, times the inverse of the mean distance to the nearest historical road of the municipalities in that same non-local area. The main results are reported in Panel B of Table 3.[29] We test if our IVs are weak by reporting the Kleibergen-Paap (2006) rk Wald F statistic – hereafter, F-statistic. As seen in Panel B, both IVs have F-statistics higher than the usual threshold of 10. They are as high as 79.8 for CIV2 (the one using the 18th-century itineraries), indicating instruments are relevant to identifying the parameters of interest. Indeed, the relevance of each CIV is reflected by the positive and significant effects on market potential, as shown by the first-stage statistics in the lower part of Table 3.[30]

**Table 3. PPML estimations: Firm births**

| Dependent variables: Count of firm births | (1) 1-year interval | (2) 5-year interval |
|---|---|---|
| **Panel A: PPML-FE results** | | |
| Market potential (*road effect*) | 0.422* | 0.586*** |
| | (0.230) | (0.130) |
| *Pseudo* $R^2$ | 0.873 | 0.921 |
| **Panel B: PPML-FEIV results** | | |
| Market potential (*road effect*) | | |
| CIV1: the fraction of MW+EW | 1.878*** | 1.228** |
| *The Roman major roads* | (0.691) | (0.480) |
| CIV2: the fraction of MW+EW | 1.597** | 1.339*** |
| *The 18th-century itineraries* | (0.662) | (0.441) |
| | First-stage results: Market potential (*road effect*) as the dependent variable | |
| CIV1 | 9.296*** | |
| | (1.179) | |
| KP F-statistic | 62.18 | |

[29] Estimated coefficients for control variables have not been reported here to save space. For a presentation and discussion of these results, please refer to Appendix D.

[30] Table C3 presents the results of a reduced-form regression of firm births on the two CIVs estimated using PPML. The reduced-form relationship between the two CIVs and our dependent variable is significant at 1% for both the 1-year and 5-year interval, providing additional evidence that weak identification does not appear to be an issue in this context (Chernozhukov & Hansen, 2008).

| | | |
|---|---|---|
| CIV2 | 1.056*** | |
| | (0.118) | |
| KP F-statistic | 79.78 | |
| Observations | 24,475 | 25,850 |

**Notes.** * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. In parentheses: Bootstrap (cluster) standard errors (300 reps.) and cluster-robust for first-stage coefficient. "1-year interval" stands for the count of establishment births measured in 1992, 2002, and 2012. "5-year interval" stands for the sum of establishment births measured in 1992-1996, 2002-2006, and 2012-2016. All regressions include control variables (urbanization and localization economies, labor cost, market size, and human capital), municipality fixed effects, and Sector of Activity (CAE) × Year fixed effects. "MW+EW" stands for motorways and expressways. "KP F-statistic" stands for the Kleibergen-Paap (2006) rk Wald F statistic. Market potential and control variables are included in logarithmic form, except for localization economies, since they contain a considerable number of zeros.

Estimates for firm births from the PPML-FEIV are now highly significant across both CIVs and time intervals. The main results also show that the point estimates are larger than PPML-FE estimates. More specifically, when examining CIV2 estimates (the IV with the highest F-statistic) and comparing columns 1 and 2 in Panel A and Panel B, we find that a 10% increase in market potential leads to a 16.0% increase in births for the 1-year interval and a 13.4% increase for the 5-year interval. Conversely, the corresponding estimates in Panel A are 4.2% and 5.9%, respectively. This significant difference between PPML-FE and PPML-FEIV estimates may indicate that the former is biased (downwards) due to endogeneity. In recent studies, Rocha et al. (2023, 2024) conducted a cross-sectional analysis to assess the effect of motorways on various aspects of local economic dynamics (e.g., population, employment, business turnover, and Gross Value Added) in Portugal and found a similar pattern, i.e., TSLS coefficients are significantly higher than those derived from OLS. The authors constructed their IVs based on historical maps from the 18th-century itineraries and the major roads of a 1945 road plan; therefore, they did not use maps from the Roman roads. In other words, they find evidence of the same downward bias using different methods and historical IVs and focusing on other variables.

Although comparing results is not straightforward due to differences in model specification and empirical strategy, our PPML-FEIV estimates suggest a higher effect of market potential on firm birth compared to previous studies for Portugal and Spain. For Portugal, Melo et al. (2010) find that a 10% increase in market potential increases births by 1.4% to 3.2%, depending on the economic sector. For Spain, Holl and Mariotti (2018b) find that a 10% increase in market potential results in a 1.7% (albeit not significant) increase in logistic firm births for the Poisson Regression Model, a 3.5% and 4.4% increase for the zero-inflated models, and an 11.1% increase for the Negative

Binomial Regression Model. In the present study, we find that a 10% increase in market potential leads to an increase in firm births ranging from 12.3% to 18.8%. Notably, we suspect these differences in estimated elasticities reflect our empirical approach, which addresses endogeneity by combining FE and IV methods.

Moreover, point estimates of the coefficient of interest for firm births are relatively stable across time and both IV estimations.[31] As already discussed, there is a positive and statistically significant association between market potential and firm births, ranging from 1.6 to 1.9 for the 1-year interval and 1.2 to 1.3 for the 5-year interval. This finding is interesting because we construct two CIVs based on historical IVs with distinct underlying logics. The main Roman roads connected a vast empire that ended around fifteen centuries ago, during which present-day Portugal was a peripheral territory. In contrast, the 18th-century itineraries formed a dense network of mainly precarious dirt roads used for short-distance travel. As shown in the lower part of Table 3, the F-statistics differ significantly between the two CIVs, indicating that they are considerably different in their relationship with the endogenous variable. The CIV2 appears to be stronger than CIV1, as evidenced by its higher F-statistic of 79.8 compared to 62.2 for CIV1. In sum, the fact that we use distinct sources of variation to construct our CIVs and obtain similar results is reassuring and suggests that we are indeed estimating a causal effect of market potential on the number of firm births.

### 5.2. Robustness analysis

#### 5.2.1. Changes to the non-local time-variant historical instruments

Our results suggest that increases in road accessibility, measured by market potential, positively affect firm births. In the following subsections, we assess the robustness of our findings by examining the sensitivity to changes in the construction of our non-local time-variant historical instrument. This involves modifying either the concept of non-locality or the measures of the historical IVs.

[31] Table C4 in Appendix C illustrates the stability of our results across the various sets of controls. Starting with a baseline specification that includes only fixed effects (columns 1 and 2), we progressively add controls such as agglomeration economies and labor market variables (columns 3 to 6).

*Changes to both components: enlarging the concept of non-locality*

First, we modify the CIVs described in Equation (2) by excluding not only the municipality's own contribution and that of its neighbors but also the contribution from the neighbors of neighbors. The idea is to enlarge the concept of non-locality of our CIVs. By excluding a larger set of neighbors, we expect to further remove potential sources of endogeneity due to similar characteristics in the nearby area, thereby reinforcing the likelihood of the exogeneity of our CIVs.

**Table 4. PPML-FEIV estimation: Excluding the contribution of the neighbors of neighbors**

| Dependent variables: Count of firm births | (1) 1-year interval | (2) 5-year interval |
|---|---|---|
| Market potential (*road effect*) | | |
| CIV1a: the fraction of MW+EW | 1.810** | 1.316** |
| *The Roman major roads* | (0.767) | (0.546) |
| CIV2a: the fraction of MW+EW | 1.505** | 1.354*** |
| *The 18th-century itineraries* | (0.738) | (0.517) |
| | First-stage results: Market potential (*road effect*) as the dependent variable | |
| CIV1a | 8.867*** | |
| | (1.293) | |
| KP F-statistic | 47.02 | |
| CIV2a | 0.974*** | |
| | (0.135) | |
| KP F-statistic | 51.78 | |
| Observations | 24,475 | 25,850 |

**Notes.** * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. In parentheses: Bootstrap (cluster) standard errors (300 reps.) and cluster-robust for first-stage coefficient. "1-year interval" stands for the count of establishment births measured in 1992, 2002, and 2012. "5-year interval" stands for the sum of establishment births measured in 1992-1996, 2002-2006, and 2012-2016. The suffix "a" after CIV indicates the exclusion of contributions from the neighbors of neighbors. All regressions include control variables (urbanization and localization economies, labor cost, market size, and human capital), municipality fixed effects, and Sector of Activity (CAE) × Year fixed effects. "MW+EW" stands for motorways and expressways. "KP F-statistic" stands for the Kleibergen-Paap (2006) rk Wald F statistic. Market potential and control variables are included in logarithmic form, except for localization economies, since they contain a considerable number of zeros.

On the one hand, additionally excluding the neighbors of neighbors apparently justifies our strategy to propose a non-local area to construct our CIVs. The estimates in Table 4 are very similar to the elasticities in Table 3, with coefficients being statistically significant at the 5% level, except for the estimates regarding the 5-year interval, which are significant at the 1% level. Columns 1 and 2 in Table 4 report, respectively, elasticities that range from 1.3 to 1.8, which are close to the range of 1.2 to 1.9 in Table 3. While one might argue that this outcome was expected given the high pairwise correlation between the CIVs used in both tables (about 0.95, as shown in Table C2), it remains a noteworthy finding. It suggests that by excluding the municipality's own contribution to the road network and those of its neighbors, we effectively remove local influences, as the results remain stable and statistically significant despite the additional exclusion of local contributions. On the other hand, precisely because of this further exclusion of local similar characteristics, the relevance of the IVs decreases as the relationship between the IVs and market potential weakens. This explains the lower F-statistics shown in the lower part of Table 4, although they are still well above the conventional threshold of 10, with values of 47.0 for the CIV1a and 51.8 for CIV2a.

*Changes to the time-invariant component: the length of the historical network.*

We now adapt the historical component of our CIV by using the total length of the historical roads instead of the inverse of the straight-line mean distance from the municipality centroid to the nearest historical road. The length of the historical network is aggregated within each NUTS2 region, excluding the municipality's own contribution and that of its neighbors. The total length of the historical roads is expected to affect market potential positively. Therefore, the sign when using either inverse distance or total length of historical roads is expected to be positive – as is indeed the case, as shown in the lower part of Table 5.

**Table 5. PPML-FEIV estimation: Using the total length of the historical road**

| Dependent variables:<br>Count of firm births | (1)<br>1-year interval | (2)<br>5-year interval |
|---|---|---|
| Market potential (*road effect*) | | |
| CIV1b: the fraction of MW+EW | 1.887*** | 1.176*** |
| *The Roman major roads* | (0.625) | (0.452) |
| CIV2b: the fraction of MW+EW | 0.969 | 0.955** |

| | | |
|---|---|---|
| *The 18th-century itineraries* | (0.605) | (0.406) |
| | First-stage results: Market potential (*road effect*) as the dependent variable | |
| CIV1b | 0.0015*** | |
| | (0.0002) | |
| KP F-statistic | 77.06 | |
| CIV2b | 0.0002*** | |
| | (0.00002) | |
| KP F-statistic | 97.74 | |
| Observations | 24,475 | 25,850 |

**Notes.** * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. In parentheses: Bootstrap (cluster) standard errors (300 reps.) and cluster-robust for first-stage coefficient. "1-year interval" stands for the count of establishment births measured in 1992, 2002, and 2012. "5-year interval" stands for the sum of establishment births measured in 1992-1996, 2002-2006, and 2012-2016. The suffix "b" after CIV indicates the use of the total length of the historical road. All regressions include control variables (urbanization and localization economies, labor cost, market size, and human capital), municipality fixed effects, and Sector of Activity (CAE) × Year fixed effects. "MW+EW" stands for motorways and expressways. "KP F-statistic" stands for the Kleibergen-Paap (2006) rk Wald F statistic. Market potential and control variables are included in logarithmic form, except for localization economies, since they contain a considerable number of zeros.

As before, the results from Table 5 are similar to the main findings. According to these estimates, elasticities range from 1.0 to 1.9, similar to the elasticities in Table 3. Although the F-statistics for the CIVs are now higher compared to the main CIV in Table 3, reaching 77.1 for CIV1b and 97.4 for CIV2b, the elasticities are estimated with less precision, particularly for the CIV2b estimates for the 1-year interval. Nevertheless, the estimates remain statistically significant for CIV1b and CIV2b estimates for the 5-year interval. Overall, despite the variability in precision, using the total length of historical roads as our measure of the historical IVs yields results consistent with those obtained using the inverse of the mean distance, indicating that our main findings are robust across different specifications.

### 5.2.2. Other tests

In this section, we submit our main PPML estimates to five additional tests (we report these estimates in Table C5 to Table C9 in Appendix C to save space).[32] First, we

[32] A referee noted that a nearby motorway could provide accessibility advantages that the market potential measure may not fully capture. However, in Portugal, an important issue is that the extensive development of the motorway network was one of the primary drivers, if not the main driver, behind the significant improvements in travel times by road and, consequently, market potential across the country. In simpler terms, proximity to motorways increases market potential. Therefore, our market potential measure, particularly the *road effect*, is likely to overlap with the information provided by additional dummy variables controlling for motorway proximity. Nonetheless, to test whether this is the case, we included

test whether our claim that the variation in road accessibility largely explains the variation in market potential. We estimate the effect of market potential on firm location using the complete measure, i.e., the measure of market potential as defined in Equation (4). The results from Table C5 reiterate our initial claim that road accessibility is the main driver of the impacts of an enlarged market potential. Using the complete measure of market potential yields results that are very similar to the main findings. Elasticities range from 1.2 to 1.9, virtually the same as the corresponding elasticities in Table 3.

Secondly, we examine the impact of excluding the two metropolitan areas of Lisbon and Porto from the sample. As described in Section 4.2, these two metropolitan areas account for roughly half of the country's population and jobs, encompassing economic dynamics that might not be comparable with the rest of the country. One might thus hypothesize that excluding municipalities that belong to these two metropolitan areas may attenuate the effect of market potential on firm births as producers will have limited access to the country's most important economic hubs. However, the results reported in Table C6 show the opposite. In most cases, the size of the estimated coefficients is more than double the baseline results.[33] More specifically, considering the CIV2 estimates, a 10% increase in market potential leads to an increase in births of about 37.5% for the 1-year interval and about 36.0% for the 5-year interval. This is compared to the corresponding effect of about 15.1% and 13.5% increase in births in Table 3, respectively. Probably, as the country became more interconnected by improvements in the road network (e.g., motorways and expressways) during the 20 years of study, the incentives for entrepreneurs to open new businesses in these regions more than compensated for the lack of access to the largest economic centers in the country, as they benefited from better and faster access to nearby markets.

Third, to examine the heterogeneous effects across industries, we estimate the specifications reported in Table 3 by restricting our sample to the two economic sectors:

---

dummy variables indicating whether a municipality falls within specific distance ranges (i.e., less than 10 km, 10–20 km, 20–50 km, and over 50 km) from the nearest motorway ramp. As shown in Table C10 in Appendix C, including these dummies does not significantly alter the main findings of our analysis. Furthermore, the coefficients of the dummies are generally not statistically significant. The complete table of results is available upon request.

[33] The F-statistics for both IVs are generally lower, with values of 23.8 for CIV1 and 23.3 for CIV2. Yet, these values remain above the conventional threshold of 10.

manufacturing and services.[34] Estimates in columns 1 to 4 of Table C7 suggest that the impact of market potential is observed only in the service sector, with positive and statistically significant effects across both CIV and time periods. No significant effects are found for manufacturing. For services, a 10% increase in market potential leads to an increase in firm births of approximately 16.9% to 25.1% over the 1-year interval and 13.7% to 16.1% over the 5-year interval. These results are not surprising. Firms in the service sector are highly dependent on the location of their final consumers. Unlike manufacturing, which primarily produces transportable goods and is less constrained by the geographic location of consumers, the services sector relies heavily on face-to-face interactions. Consequently, proximity to large markets becomes a critical determinant for services, making road access to these markets an especially relevant factor influencing new business formation in this sector.

Fourth, since physical geography may have influenced the construction of ancient roads and modern transport infrastructure, we test our main results by including an interaction term between a time-invariant variable representing local geographic conditions and time-period fixed effects. We attempt to account for potential differential changes in road construction costs due to local geographic features, which could impact both ancient and modern roads (Andersson et al., 2023). The local geography measure is defined as the standard deviation of terrain elevation within municipality $j$ – a measure of terrain ruggedness.[35] As seen from Table C8, the coefficients remain stable after adding the local geography-year FE. The F-statistics for both instruments are now slightly lower, with values of 49.58 and 68.10, respectively. Elasticities are 1.5 and 1.7 for the 1-year interval and 1.3 and 1.5 for the 5-year interval. These values are very close to the corresponding elasticities reported in Table 3, which are 1.6 and 1.9 for the 1-year interval and 1.2 and 1.3 for the 5-year interval. This consistency suggests that including local geographic factors does not significantly affect our main findings, further supporting the robustness of the results.

[34] The specific economic activities within each group are detailed in Table A4 in Appendix A. The breakdown of births by sector for the 1-year and 5-year intervals, respectively, is as follows: Manufacturing accounts for 12.7% and 11.4% of the total births, and Services accounts for 66.9% and 69.7%. The remaining sectors (i.e., Primary, Construction, and Utilities) account for 20.4% and 18.9%, respectively. Figures C2 and C3 illustrate the spatial distribution of new firms in the manufacturing and services sectors for 1992, 2002, and 2012.

[35] This measure is obtained from Afonso et al. (2024).

Finally, we assess the sensitivity of our primary estimates by altering the time interval under analysis. Specifically, we extend the 5-year interval to an 8-year interval, now including the additional years 1997-1999, 2007-2009, and 2017-2019. The inclusion of 2007, 2008, and 2009 is particularly interesting due to the 2008 financial crisis. For convenience, column 1 of Table C9 replicates our main findings from column 1 of Table 3, while column 2 presents estimations for the same equation but for the extended 8-year interval mentioned above. The point estimates are only statistically significant for estimates obtained from CIV2 and, overall, lower in magnitude. This declining effect over time could be expected since we are evaluating the impact of a network variation that occurred eight years earlier or more. Indeed, as noted in Section 3.2.1, we measure the state of the road network one, five, or eight years earlier than the years when the number of firm births was observed, but the actual variation in the network may have occurred even earlier. Hence, it is reasonable to expect that the effects of road improvements will decrease over time. Nevertheless, even if we consider that the lower point estimates are due to the specific characteristics of the years included, mainly years of some economic turmoil, the estimate for CIV2 remains relatively close to the main estimates in Table 3.

None of the modifications considered to the sample or models seem to significantly influence our results, as the estimated coefficients remain broadly stable across the five tests and are very similar to those in Table 3. Exceptions are the cases where we exclude LMA and PMA municipalities, which appear to result in greater effects on firm births, and the heterogeneous analysis by industry group, where we observe that births from service industries are particularly responsive to changes in market potential.

## 6. Conclusion

This paper investigates whether the municipalities that benefited from enlarged market potential due to improved road accessibility experienced more firm births than less connected municipalities. We focus on Portugal from 1991 to 2016, during which the country experienced a remarkable transformation in its road network: on average, more than 100 km of motorways were built annually.

Municipality-level regressions using the PPML-FEIV estimator provide causal evidence that improvements in road accessibility, measured by market potential, have positively impacted private business activity, increasing the number of firm births. Estimated elasticities range from 1.6 to 1.9 for firm births measured over 1-year intervals and from 1.2 to 1.3 for firm births measured over 5-year intervals. These results are both statistically significant and economically meaningful. To illustrate through a simple exercise, given that the average market potential increased by 1.5% annually (see Table 2) and the estimated elasticities are, on average, approximately 1.6%, this could imply a 2.4% increase in firm births annually.

Our findings are robust to changes in model specification and additional robustness checks. For instance, reinforcing the non-locality of our CIVs by further excluding nearby areas and using alternative measures of historical roads does not appear to impact our main conclusions substantially. Interestingly, when we test our results by excluding municipalities from the LMA and PMA, we observe a greater effect on firm births. This finding might be explained by the non-local nature of market potential, suggesting positive spillover effects in the sense that new motorways and expressways increased overall road accessibility, thus enhancing the attractiveness of municipalities that, despite not being direct targets of major investments in road infrastructure or lacking access to the two largest markets, benefit from these new opportunities created by shorter road distances and improved connections between all other municipalities in the country. Another important finding emerges when analyzing heterogeneous effects across industries: the impact of market potential is primarily concentrated in service industries. These industries, which rely heavily on proximity to final consumers, are particularly responsive to changes in market potential, resulting in stronger effects on firm births. A plausible explanation is that improved accessibility facilitates access to larger markets and a broader pool of customers, alleviating constraints imposed by geographic location. This, in turn, has provided strong incentives for entrepreneurs to establish new businesses within this sector.

From a policy perspective, these findings are important to urban and regional development. Large-scale road investments can stimulate local long-run economic growth, but, as a side effect, they can also increase regional disparities. As much of the funding for road improvements in Portugal came from European structural funds aimed at reducing regional disparities within and between EU member states, policymakers should carefully consider the potential for spatial concentration effects due to transport improvements, which may increase regional disparities rather than reduce them. Our findings suggest that constructing motorways and expressways, measured by market potential, positively affected economic activity through increased firm births. The extent to which the estimated positive average local impact represents genuine growth or displacement effects remains a matter of great importance but an issue for future research.

## References


Afonso, N., Melo, P., Rocha, B. T., & Abreu, J. (2024). TiTuSS Transport Database (Version 5) [Data set]. *Mendeley Data*. https://doi.org/10.17632/ry5dkty7t7.5

Afonso, N., Melo, P. C., Rocha, B. T., & de Abreu e Silva, J. (2025). Going big in a small country: Fifty years of motorways in Portugal. *Revista Portuguesa de Estudos Regionais, 70*(1).

Agrawal, A., Galasso, A., & Oettl, A. (2017). Roads and innovation. *Review of Economics and Statistics*, *99*(3), 417–434.

Alañón-Pardo, Á., & Arauzo-Carod, J.-M. (2013). Agglomeration, accessibility, and industrial location: Evidence from Spain. *Entrepreneurship & Regional Development*, *25*(3–4), 135–173.

Alañón-Pardo, A., Walsh, P. J., & Myro, R. (2018). Do neighboring municipalities matter in industrial location decisions? Empirical evidence from Spain. *Empirical Economics*, *55*(3), 1145-1179.

Andersson, D., Berger, T., & Prawitz, E. (2023). Making a market: Infrastructure, integration, and the rise of innovation. *Review of Economics and Statistics*, *105*(2), 258–274.

Anyadike-Danes, M., & Hart, M. (2018). All grown up? The fate after 15 years of a quarter of a million UK firms born in 1998. *Journal of Evolutionary Economics*, *28*, 45-76.

Arauzo-Carod, J. M., & Manjón-Antolín, M. (2012). (Optimal) spatial aggregation in the determinants of industrial location. *Small Business Economics*, *39*(3), 645-658.

Arauzo-Carod, J.-M., Liviano-Solis, D., & Manjón-Antolín, M. (2010). Empirical Studies in Industrial Location: An Assessment of Their Methods and Results. *Journal of Regional Science*, *50*(3), 685–711.

Audretsch, D. B., Dohse, D. C., & dos Santos, J. P. (2020). The effects of highway tolls on private business activity—results from a natural experiment. *Journal of Economic Geography*, *20*(6), 1331-1357.

Balbontin, C., & Hensher, D. A. (2019). Firm-specific and location-specific drivers of business location and relocation decisions. *Transport Reviews*, *39*(5), 569-588.

Baum-Snow, N. (2007). Did highways cause suburbanization? *The quarterly journal of economics*, *122*(2), 775–805.

Baum-Snow, N., Brandt, L., Henderson, J. V., Turner, M. A., & Zhang, Q. (2017). Roads, railroads, and decentralization of Chinese cities. *Review of Economics and Statistics*, *99*(3), 435-448.

Beine, M., Bertoli, S., & Fernández-Huertas Moraga, J. (2016). A practitioners' guide to gravity models of international migration. *The World Economy*, *39*(4), 496-512.

Bhat, C. R., Paleti, R., & Singh, P. (2014). A spatial multivariate count model for firm location decisions. *Journal of Regional Science*, *54*(3), 462–502.

Blackburn, M. L. (2015). The relative performance of Poisson and negative binomial regression estimators. *Oxford Bulletin of Economics and Statistics*, *77*(4), 605–616.

Bottasso, A., Conti, M., Ferrari, C., & Tei, A. (2022). Roads to innovation: Evidence from Italy. *Journal of Regional Science, 62*(5), 1115–1138.

Branco, C., Dohse, D. C., dos Santos, J. P., & Tavares, J. (2023). Nobody's gonna slow me down? The effects of a transportation cost shock on firm performance and behavior. *Journal of Urban Economics*, *136*, 103569.

Brughmans, T., De Soto, P., Pažout, P., & Bjerregaard Vahlstrup, J. (2024). *Itiner-e: The digital atlas of ancient roads*. Retrieved from https://itiner-e.org/

Capozza, C., Salomone, S., & Somma, E. (2018). Local industrial structure, agglomeration economies and the creation of innovative start-ups: evidence from the Italian case. *Entrepreneurship & Regional Development*, *30*(7-8), 749-775.

Carpenter, C. W., Van Sandt, A., & Loveridge, S. (2021). Empirical methods in business location research. *Regional Studies, Regional Science*, *8*(1), 344–361.

Chernozhukov, V., & Hansen, C. (2008). Instrumental variable quantile regression: A robust inference approach. *Journal of Econometrics*, *142*(1), 379-398.

Coad, A., Daunfeldt, S. O., & Halvarsson, D. (2018). Bursting into life: firm growth and growth persistence by age. *Small Business Economics*, *50*(1), 55-75.

Coll-Martínez, E., Jové-Llopis, E., & Teruel, M. (2022). The city of start-ups: Location determinants of start-ups in emergent industries in Barcelona. *Growth and Change*.

Correia, S., Guimarães, P., & Zylkin, T. (2020). Fast Poisson estimation with high-dimensional fixed effects. *The Stata Journal*, *20*(1), 95-115.

Coughlin, C. C., & Segev, E. (2000). Location determinants of new foreign-owned manufacturing plants. *Journal of Regional Science*, *40*(2), 323–351.

Dalgaard, C. J., Kaarsen, N., Olsson, O., & Selaya, P. (2022). Roman roads to prosperity: Persistence and non-persistence of public goods provision. *Journal of Comparative Economics, 50*(4), 896–916.

Daunfeldt, S. O., Elert, N., & Rudholm, N. (2013). Start-ups and firm in-migration: evidence from the Swedish wholesale industry. *The annals of regional science*, *51*(2), 479–494.

De Benedictis, L., Licio, V., & Pinna, A. M. (2023). From the historical Roman road network to modern infrastructure in Italy. *Journal of Regional Science*, *63*(5), 1162-1191.

De Soto, P. (2019). Network analysis to model and analyse Roman transport and mobility. *Finding the Limits of the Limes: Modelling Demography, Economy, and Transport on the Edge of the Roman Empire*, 271–289.

De Soto, P., & Carreras, C. (2021). The role of politics in the historical transport networks of the Iberian Peninsula and the Pyrenees from Roman times to the nineteenth century. *Social Science History*, *45*(2), 233-260.

Duranton, G., & Turner, M. A. (2011). The fundamental law of road congestion: Evidence from US cities. *American Economic Review*, *101*(6), 2616-52.

Duranton, G., & Turner, M. A. (2012). Urban growth and transportation. *Review of Economic Studies*, *79*(4), 1407-1440.

Eurostat-OECD. (2007). *Eurostat-OECD manual on business demography statistics*. Paris: OECD.

Figueiredo, O., Guimaraes, P., & Woodward, D. (2002). Home-field advantage: location decisions of Portuguese entrepreneurs. *Journal of Urban Economics*, *52*(2), 341-361.

Flueckiger, M., Hornung, E., Larch, M., Ludwig, M., & Mees, A. (2022). Roman transport network connectivity and economic integration. *The Review of Economic Studies, 89*, 774–810.

Garcia-López, M. À. (2019). All roads lead to Rome… and to sprawl? Evidence from European cities. *Regional Science and Urban Economics*, *79*, 103467.

Garcia-López, M. À., Holl, A., & Viladecans-Marsal, E. (2015). Suburbanization and highways in Spain when the Romans and the Bourbons still shape its cities. *Journal of Urban Economics*, *85*, 52-67.

Ghani, E., Goswami, A. G., & Kerr, W. R. (2016). Highway to success: The impact of the Golden Quadrilateral Project for the location and performance of Indian manufacturing. *The Economic Journal, 126*(591), 317–357.

Gibbons, S., Heblich, S., Pinchbeck, E.W. (2024). The spatial impacts of a massive rail disinvestment program: The Beeching Axe. *Journal of Urban Economics*, 143, 10369.

Gibbons, S., Lyytikäinen, T., Overman, H. G., & Sanchis-Guarner, R. (2019). New road infrastructure: the effects on firms. *Journal of Urban Economics*, *110*, 35-50.

Graham, D. J. (2007). Agglomeration, productivity and transport investment. *Journal of transport economics and policy,* 41(3), 317-343.

Graham, D. J., & Gibbons, S. (2019). Quantifying Wider Economic Impacts of agglomeration for transport appraisal: Existing evidence and future directions. *Economics of Transportation*, *19*, 100–121.

Graham, D. J., & Van Dender, K. (2011). Estimating the agglomeration benefits of transport investments: some tests for stability. *Transportation*, *38*(3), 409-426.

Guimarães, P. (2008). The fixed effects negative binomial model revisited. *Economics Letters*, *99*(1), 63-66.

Guimarães, P., Figueiredo, O., & Woodward, D. (2000). Agglomeration and the Location of Foreign Direct Investment in Portugal. *Journal of Urban Economics*, *47*(1), 115–135.

Guimarães, P., Figueirdo, O., & Woodward, D. (2003). A Tractable Approach to the Firm Location Decision Problem. *Review of Economics and Statistics*, *85*(1), 201–204.

Guimarães, P., Figueiredo, O., & Woodward, D. (2004). Industrial Location Modeling: Extending the Random Utility Framework. *Journal of Regional Science*, *44*(1), 1–20.

Harris, C. D. (1954). The Market as a Factor in the Localization of Industry in the United States. *Annals of the Association of American Geographers*, *44*(4), 315–348.

Heidt, P., & Kasim, M. T. (2020). The effects of highways on school segregation. *Papers in Regional Science*, *99*(5), 1261-1281.

Hidalgo, A., Riccaboni, M., & Velázquez, F. J. (2024). The effect of short-term rentals on local consumption amenities: Evidence from Madrid. *Journal of Regional Science*.

Holl, A. (2004a). Manufacturing location and impacts of road transport infrastructure: Empirical evidence from Spain. *Regional Science and Urban Economics*, 34(3), 341–363.

Holl, A. (2004b). Start-ups and relocations: Manufacturing establishment location in Portugal. *Papers in Regional Science*, 83(4), 649–668.

Holl, A. (2004c). Transport Infrastructure, Agglomeration Economies, and Firm Birth: Empirical Evidence from Portugal. *Journal of Regional Science*, 44(4), 693–712.

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

Holl, A. (2016). Highways and productivity in manufacturing firms. *Journal of Urban Economics*, *93*, 131–151.

Holl, A., & Mariotti, I. (2018a). Highways and firm performance in the logistics industry. *Journal of Transport Geography*, *72*, 139-150.

Holl, A., & Mariotti, I. (2018b). The geography of logistics firm location: the role of accessibility. *Networks and Spatial Economics*, *18*(2), 337-361.

Ignatov, A. (2023). European Highway networks, transportation costs, and regional income. *Regional Science and Urban Economics*, 103969.

Jofre-Monseny, J., Marín-López, R., & Viladecans-Marsal, E. (2014). The determinants of localization and urbanization economies: Evidence from the location of new firms in Spain. *Journal of Regional Science*, *54*(2), 313-337.

Kim, H., Ahn, S., & Ulfarsson, G. F. (2018). Transportation infrastructure investment and the location of new manufacturing around South Korea's West Coast Expressway. *Transport Policy*, *66*, 146-154.

Kleibergen, F., & Paap, R. (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics*, *133*(1), 97-126.

Levkovich, O., Rouwendal, J., & van Ommeren, J. (2020). The impact of highways on population redistribution: the role of land development restrictions. *Journal of Economic Geography*, *20*(3), 783-808.

Lin, W., & Wooldridge, J. M. (2019). Testing and correcting for endogeneity in nonlinear unobserved effects models. In *Panel data econometrics* (pp. 21-43). Academic Press.

Liviano, D., & Arauzo-Carod, J.-M. (2013). Industrial location and interpretation of zero counts. *The Annals of Regional Science*, *50*(2), 515–534.

Liviano, D., & Arauzo-Carod, J.-M. (2014). Industrial Location and Spatial Dependence: An Empirical Application. *Regional Studies*, *48*(4), 727 743.

Manjón-Antolín, M. C., & Arauzo-Carod, J. M. (2011). Locations and relocations: determinants, modelling, and interrelations. *The Annals of Regional Science*, *47*(1), 131-146.

Maré, D. C., & Graham, D. J. (2013). Agglomeration elasticities and firm heterogeneity. *Journal of Urban Economics*, 75, 44-56.

Marshall, A. (1890). *Principles of Economics*. The Online Library of Liberty. Palgrave Macmillan: London. https://doi.org/10.1057/9781137375261

Matos, A. T. (1980). *Transportes e Comunicação em Portugal, Açores e Madeira (1750–1850)*. Universidade dos Açores.

McCormick, M., Huang, G., Zambotti, G., & Lavash, J. (2013). *Roman road network (version 2008)*. DARMC Scholarly Data Series, Data Contribution Series, 2013-5.

Mejia-Dorantes, L., Paez, A., & Vassallo, J. M. (2012). Transportation infrastructure impacts on firm location: The effect of a new metro line in the suburbs of Madrid. *Journal of Transport Geography*, *22*, 236-250.

Melo, P. C., Graham, D. J., & Noland, R. B. (2010). Impact of Transport Infrastructure on Firm Formation: Evidence from Portuguese Municipalities. *Transportation Research Record: Journal of the Transportation Research Board*, 2163(1), 133–143.

Miroudot, S., & Rigo, D. (2022). Multinational production and investment provisions in preferential trade agreements. *Journal of Economic Geography*, *22*(6), 1275-1308.

Moeller, K. (2018). Culturally clustered or in the cloud? How amenities drive firm location decision in Berlin. *Journal of Regional Science*, *58*(4), 728–758.

Möller, J., & Zierer, M. (2018). Autobahns and jobs: A regional study using historical instrumental variables. *Journal of Urban Economics*, *103*, 18-33.

Mota, I., & Brandão, A. (2013). The determinants of location choice: Single establishments versus multi-establishments. *Papers in Regional Science*, 92(1), 31-49.

Nilsson, I. M., & Smirnov, O. A. (2016). Measuring the effect of transportation infrastructure on retail firm co-location patterns. *Journal of Transport Geography*, *51*, 110-118.

Nocito, S., Sartarelli, M., & Sobbrio, F. (2023). A beam of light: Media, tourism and economic development. *Journal of Urban Economics*, *137*, 103575.

Otsuka, A. (2008). Determinants of new firm formation in Japan: A comparison of the manufacturing and service sectors. *Economics Bulletin*, *18*(4), 1–7.

Pacheco, E. (2004). *Evolução e Planeamento da Rede de Transportes em Portugal. Alteração das Acessibilidades e Dinâmicas Territoriais da Região Norte* (Doctoral dissertation, PhD thesis, Faculdade de Letras, Universidade do Porto, Porto).

Percoco, M. (2016). Highways, local economic structure and urban development. *Journal of Economic Geography, 16*(5), 1035–1054.

Pereira, A. M., & Pereira, R. M. (2018). Are all infrastructure investments created equal? The case of Portugal. *Journal of Infrastructure, Policy and Development*, 2(1), 67-86.

Redding, S. J., & Turner, M. A. (2015). Transportation costs and the spatial organization of economic activity. In *Handbook of Regional and Urban Economics* (Vol. 5, pp. 1339–1398). Elsevier.

Roca, J. D. L., & Puga, D. (2017). Learning by working in big cities. *The Review of Economic Studies, 84*(1), 106–142.

Rocha, B. T., Afonso, N., Melo, P. C., & e Silva, J. D. A. (2022). What drives the allocation of motorways? Evidence from Portugal's fast-expanding network. *Journal of Transport Geography*, *102*, 103384.

Rocha, B. T., Melo, P. C., Afonso, N., & e Silva, J. A. (2023). The local impacts of building a large motorway network: Urban growth, suburbanisation, and agglomeration—Economics *of Transportation, 34*, 100302.

Rocha, B. T., Melo, P. C., Afonso, N., & de Abreu e Silva, J. (2024). Motorways, local economic activity and commuting. *Regional Studies*, *58*(1), 164-177.

Rosselló-Nadal, J., & Santana-Gallego, M. (2024). Toward a smaller world. The distance puzzle and international border for tourism. *Journal of Transport Geography*, *115*, 103809.

Rupasingha, A., & Marré, A. W. (2020). Moving to the hinterlands: agglomeration, search costs and urban to rural business migration. *Journal of Economic Geography*, *20*(1), 123-153.

Santagata, G. (2022). Roads and intra-national trade: Evidence from Italian regions. *Papers in Regional Science, 101*(1), 1–23.

Santos Silva, J. M. C., & Tenreyro, S. (2006). The log of gravity. *The Review of Economics and Statistics*, *88*(4), 641-658.

Santos Silva, J. M. C., & Tenreyro, S. (2011). Further simulation evidence on the performance of the Poisson pseudo-maximum likelihood estimator. *Economics Letters*, *112*(2), 220-222.

Santos Silva, J. M. C., & Tenreyro, S. (2022). The log of gravity at 15. *Portuguese Economic Journal*, 1-15.

Schlegel, T., & Backes-Gellner, U. (2023). The role of fields of study for the effects of higher education institutions on regional firm location. *Small Business Economics*, 1-70.

Von Thünen, J. H. (1842). *Der isolirte staat in beziehung auf landwirthschaft und nationalökonomie*. Leopord: Rostock [trans. Wartenberg. CM.(1966), Von Thünen's Isolated State. Pergamon. Oxford, UK].

Wahl, F. (2017). Does European development have Roman roots? Evidence from the German limes. *Journal of Economic Growth, 22*(3), 313–349.

Weber, A. (1909) Uber den Standort der Industrien, Tübingen. Tübingen: J. C. B. Mohr. University of Chicago Press, Chicago (English translation by C. Friedrich Theory of the Location of Industries.

Weidner, M., & Zylkin, T. (2021). Bias and consistency in three-way gravity models. *Journal of International Economics*, *132*, 103513.

Wooldridge, J. M. (1999). Distribution-free estimation of some nonlinear panel data models. *Journal of Econometrics*, *90*(1), 77-97.

**Appendix A**

*Harmonization of the Classification of Economic Activities (CAE)*

This appendix presents the procedure used to establish compatibility between the classification of economic activities (CAE) produced by Statistics Portugal (INE), namely: CAE-REV.1 implemented in 1973; CAE-REV.2 implemented in 1993; CAE-REV.2.1 implemented in 2003; and CAE REV.3 implemented in 2007.[1] To proceed accordingly, it was required to match CAE codes for the following periods: CAE-REV.1 from 1991 to 1994; CAE-REV.2 from 1995 to 2005; CAE-REV.2.1 from 2006 to 2007; and CAE-REV.3 from 2008 to 2019. To obtain results comparable to those in the literature and to ensure data consistency, we chose to use the CAE-REV.2.1 classification. Another reason for this decision is the fact that the revisions from REV.2.1 to REV.3 were quite drastic as they reflect technological advances and new activities developed in the late 20th century and early 21st century and, as a result, much information would be lost if the harmonization from REV.2.1 to REV.3 was carried on. Hence, we set the CAE-REV.2.1 as our baseline classification.

The procedure is as follows. The harmonization from REV.1 to REV.2, and then from Rev 2 to REV.2.1, is done in a bottom-up way. In other words, we start with the most disaggregated level from the oldest classification (REV.1 or REV.2) and compare it to the same hierarchical level in the most recent classification (Rev.2 or Rev.2.1). When this approach is not feasible, we go up one level in the classification hierarchy for which there is less detail about industries. We repeat this procedure starting from 5 digits (6 digits when available) up to the highest aggregate level – usually a letter group corresponding to macro sectors (e.g., Fishing, Construction, Manufacturing, etc.). We excluded economic activities for which it was not possible to apply this algorithm. Regarding the harmonization between REV.3 and REV.2.1, it is possible to define a function that achieves a precise one-to-one harmonization; that is, any group from REV.3 can be directly converted to a REV.2.1 group.

Using this procedure, we have identified 33 sectors in CAE-REV.1 and nine sectors in CAE-REV.2 for which it was not possible to apply the harmonization

[1] See INE (2007), *Classificação Portuguesa das Actividades Económicas Rev.3*. INE, Lisboa.

algorithm. Table A1 shows these sectors regarding CAE-REV.1 (first column) and CAE-REV.2 (second column). The observations excluded from CAE-REV.1 represent 4.3% of the total number of observations for the years 1991 to 1994.[2] As for CAE-REV.2, the excluded observations represent 0.6% of the total number of observations from 1995 to 2005.

**Table A1. Excluded sectors from CAE-REV.1 and CAE-REV.2**

| CAE-REV.1 | CAE-REV.2 |
|---|---|
| 331929, 332030, 351140, 352400, 352920, 352920, 354090, 355120, 355990, 356000, 371010, 371010, 371010, 381990, 382100, 382510, 383200, 384330, 390970, 711100, 712330, 713200, 719120, 832900, 920000, 933060, 934200, 934200, 941110, 941210, 941520, 941590, 951900, 951900, 953000, 959990 | 01420, 14121, 18240, 18240, 18240, 33202, 36210, 40102, 40202, 41000, 74810, 92720 |

Source: Author's compilation based on *Quadros de Pessoal*.

In cases where the bottom-up algorithm was unfeasible, we chose to use correspondence in an *ad hoc* manner. Table A2 shows the original and new codes, respectively.

**Table A2.Ad hoc correspondence between CAE-REV.1 and CAE-REV.2**

| Original code | CAE-REV.2 code |
|---|---|
| 111000 | 01 |
| 321910 | 17544 |
| 321990 | 17 |
| 352990 | 24 |
| 369950 | 26 |
| 371090 | 27 |
| 372090 | 27 |
| 934100 | 85324 |
| 384990 | 35 |

Source: Author's compilation based on *Quadros de Pessoal*.

The final list of 32 industries used in the analyses is presented in Table A3 below.

[2] Unexpectedly, some 6-digit codes do not exist in the official CAE table, although they do exist in the *Quadros de Pessoal* database. Therefore, observations associated with these codes were also eliminated. These observations represent only 0.1% of total observations from 1991 to 1994, which is reassuring.

## Table A3. Final list of economic sectors

| Classification similar to the literature | CAE-REV.2.1 code |
|---|---|
| Computer and related activities (IT services) | 72300, 72400, 72100, 722, 72500, 72220, 72600, 72210 |
| Construction | 45211, 45212, 45330, 45310, 45430, 45230, 45110, 45450, 45420, 45410, 45340, 45250, 45500, 45320, 45440, 45240, 45120, 45220 |
| Electricity, gas, and water | 41000, 40302, 40110, 40220, 40130, 40301, 40120, 40210 |
| Finance & Insurance | 67200, 67130, 66012, 660, 66030, 65224, 65221, 65120, J, 67120, 6512, 65110, 65230, 66020, 65210, 66011, 652, 65222, 65223, 67110, 671 |
| Hotels and restaurants | 55306, 55402, 55301, 55302, 5540, 55404, 551, 55401, 55520, 55304, 55403, 55121, 55111, 55122, 55124, 55406, 55233, 55113, 55112, 55510, 55405, 55119, 55303, 55116, 55305, 55115, 55234, 55220, 55114, 55123, 55232, 55117, 55231, 55118, 55210 |
| Land, water, air transport, and supporting services | 60240, 60230, 63110, 60220, 602, 60211, 60212, 63230, 63401, 63402, 63300, 63220, 63210, 62100, 63122, 60300, 61102, 61101, 61200, 60100, 6312, 63121, 6110, 62200, 62300 |
| Manufacture of basic metals and fabricated metal | 287, 28210, 28751, 28622, 28752, 28630, 28520, 28120, 28110, 27340, 28741, 28720, 28401, 27540, 28730, 28610, 27450, 28743, 27420, 28621, 27440, 28510, 274, 27100, 27430, 28710, 27530, 28742, 27510, 28623, 27410, 27320, 27210, 27520, 27220, 28220, 28300, 27310, 28402, 272, 27330 |
| Manufacture of chemical and chemical products | 24301, 24520, 241, 24141, 24410, 24663, 24620, 24160, 244, 24422, 24511, 24662, 24421, 24200, 24151, 24130, 24143, 24512, 24152, 24302, 24700, 24661, 24110, 24640, 24142, 24630, 24170, 2414, 24120, 24610, 2415, 24650 |
| Manufacture of coke, refined petroleum products, and nuclear fuel | 23300, 23100, 23200 |
| Manufacture of electrical machinery and apparatus n.e.c | 31620, 31100, 31400, 31500, 31202, 31300, 31610, 31201 |
| Manufacture of food products and beverages | 15820, 15860, 15932, 15811, 15510, 151, 15130, 15812, 157, 15931, 15334, 15611, 15982, 15893, 15842, 15950, 1533, 15335, 15911, 15612, 1598, 15981, 15203, 15110, 15420, 154, 15913, 15120, 1520, 15204, 15870, 15202, 15710, 15320, 15520, 15841, 15960, 15412, 15850, 15333, 15413, 15912, 15620, 15201, 15430, 15720, 15613, 153, 15331, 15892, 15920, 15880, 15830, 15411, 15891, 15332, 15310, 15970, 15940 |
| Manufacture of leather and leather products | 19301, 1930, 19200, 1910, 19101, 19302, 19102 |
| Manufacture of machinery and equipment n.e.c. | 294, 29420, 29243, 29564, 29320, 29563, 29230, 29520, 29130, 29310, 29540, 29120, 29710, 29222, 29221, 29242, 29430, 29410, 29720, 29550, 29530, 29110, 29510, 2960, 29601, 29241, 29561, 29210, 29140, 29562, 29602 |
| Manufacture of medical, precision & optical instruments, watches & clocks | 3310, 33101, 33500, 33102, 33203, 33300, 33401, 3340, 33201, 33403, 33402 |
| Manufacture of motor vehicles and transport equipment | 35430, 35420, 34200, 351, 35112, 34300, 35111, 354, 35410, 34100, 35500, 35200, 35300, 35120 |
| Manufacture of office machinery and computers | 30020, 30010 |
| Manufacture of other non-metallic mineral products | 266, 26610, 26120, 261, 26150, 26701, 2640, 26401, 26660, 26403, 26211, 26402, 26260, 26240, 26510, 26301, 26213, 26703, 26702, 26110, 26522, 26212, 2682, 26630, 26132, 26810, 26220, 26302, 26650, 26822, 26140, 26530, 26640, 26521, 26620, 26230, 26250, 26131, 26821 |
| Manufacture of pulp, paper, and paper products | 21250, 21230, 2121, 21212, 212, 21120, 21211, 21110, 21220, 21240 |

**Table A3. Final list of economic sectors** (*continuation*)

| | |
|---|---|
| Manufacture of radio, television, and communication equipment | 32300, 32100, 32200 |
| Manufacture of rubber and plastic products | 25240, 25210, 25120, 25220, 25130, 25110, 25230 |
| Manufacture of textiles, wearing apparel, dying and dressing of fur | 18221, 18222, 17110, 17510, 17400, 17230, 17240, 17210, 17710, 17720, 17544, 17220, 18230, 17542, 17600, 18100, 17301, 18302, 17250, 17303, 18210, 17120, 1830, 18301, 17130, 17522, 17150, 17541, 17302, 17170, 17140, 17521, 17160, 171, 17543, 17530, 18240 |
| Manufacture of tobacco products | 16000 |
| Manufacture of wood and wood products | 20101, 20522, 2030, 20302, 20511, 20201, 20301, 20400, 20512, 20203, 20521, 20102, 20202 |
| Other business and consultancy activities | 74140, 74872, 74850, 74700, 74500, 74202, 74150, 74110, 74120, 74201, 7487, 74300, 74401, 74820, 74600, 74860, 748, 74402, 74871, 74130, 74810 |
| Post and telecommunications | 64110, 64200, 64120 |
| Primary industries (including mining and extraction) | 02012, 14501, 01112, 14111, 01230, 01502, 01132, 01131, 14210, 01240, 14220, 14504, 01300, 14112, 01410, 01210, 01120, 01252, 014, 11100, 05011, 050, 14401, 020, 01111, 01501, 02020, 01133, 02011, 01220, C, 14121, 05020, 05013, 13205, 13203, 13202, 10300, 10101, 13100, 12000, 14130, 14502, 14122, 14403, 13204, 14301, 13201, 14402, 14503, 01420, 14302, 01251, 01134, B, 05012, 11200, 0150, 10102 |
| Publishing, printing, and reproduction of recorded media | 222, 22220, 22240, 22210, 22110, 221, 22120, 22250, 22230, 22130, 22150, 22140, 22310, 22320, 22330 |
| Real estate activities | 70310, 70320, 70120, 70110, 70200 |
| Recreational, cultural, and sporting activities | 92342, 92710, 92130, 926, 92620, 92312, 92610, 92311, 92111, 92720, 92320, 92200, 925, 92520, 92120, 92112, 92530, 92330, 92341, 92510, 92400 |
| Renting of machinery and equipment without operator and of personal and household goods | 71210, 71100, 71310, 713, 71320, 71340, 71400, 71330, 71220, 71230 |
| Research and Development | 73100, 73200 |
| Wholesale and retail trades | 51140, 51900, 51870, 52112, 51110, 51532, 52421, 52441, 51540, 518, 52488, 52485, 52410, 52472, 51475, 51520, 50200, 52120, 52220, 51180, 51460, 51190, 51341, 51370, 52463, 52483, 52482, 51390, 51474, 52422, 51550, 52442, 52487, 50500, 50300, 51422, 52111, 52443, G, 51382, 51430, 52451, 52461, 52272, 51212, 51381, 52471, 51410, 50100, 52720, 52481, 52500, 51160, 52240, 51441, 51810, 51211, 52630, 52432, 52330, 52431, 51563, 52452, 51450, 52230, 52320, 52260, 512, 50401, 52486, 51471, 51311, 52210, 52250, 51510, 51130, 5211, 51312, 51442, 52310, 52444, 52462, 51170, 51880, 51331, 51320, 51361, 51362, 51421, 51332, 51531, 51150, 51473, 52710, 51572, 51561, 51571, 50402, 51350, 51120, 52740, 51342, 51240, 52484, 51250, 52621, 51830, 52730, 51472, 51562, 51573, 51820, 51860, 51230, 51850, 52622, 51220, 51840, 52271, 52610, 52623 |

Source: Author's compilation based on *Quadros de Pessoal*.

*Economic sectors: Manufacturing and Services*

**Table A4. List of economic sectors classified as Manufacturing or Services**

| Economic Sectors | Classification similar to the literature |
|---|---|
| Manufacturing | Manufacture of basic metals and fabricated metal |
| | Manufacture of chemical and chemical products |
| | Manufacture of coke, refined petroleum products, and nuclear fuel |
| | Manufacture of electrical machinery and apparatus n.e.c |
| | Manufacture of food products and beverages |
| | Manufacture of leather and leather products |
| | Manufacture of machinery and equipment n.e.c. |
| | Manufacture of medical, precision & optical instruments,watches & clocks |
| | Manufacture of motor vehicles and transport equipment |
| | Manufacture of office machinery and computers |
| | Manufacture of other non-metallic mineral products |
| | Manufacture of pulp, paper, and paper products |
| | Manufacture of radio, television, and communication equipment |
| | Manufacture of rubber and plastic products |
| | Manufacture of textiles, wearing apparel, dying and dressing of fur |
| | Manufacture of wood and wood products |
| | Publishing, printing, and reproduction of recorded media |
| Services | Computer and related activities (IT services) |
| | Finance & Insurance |
| | Hotels and restaurants |
| | Land, water, air transport, and supporting services |
| | Other business and consultancy activities |
| | Post and telecommunications |
| | Real estate activities |
| | Recreational, cultural, and sporting activities |
| | Research and Development |
| | Wholesale and retail trades |

Source: Author's compilation based on *Quadros de Pessoal*.

## Appendix B

*Shift-share analysis: Market potential*

To simplify matters, let us admit the proof considering market potential regarding a single municipality, where we denote it by $j$. As described in Equation (4), $MP_j^t$ consists of the summation of the population ($POP_k$) in each neighboring municipality $k$ after discounting the travel time ($d_{jk}^t$):

$$MP_j^t = \sum_k \frac{POP_k^t}{d_{jk}^t} = \frac{POP_1^t}{d_{j1}^t} + \frac{POP_2^t}{d_{j2}^t} + \cdots + \frac{POP_N^t}{d_{jN}^t} \tag{8}$$

$$k = 1, \dots, j-1, j+1, \dots, N$$

where $N$ refers to all the 275 municipalities in mainland Portugal. Thus, the relative variation (growth rate) between 2001 and 1991 can be written as:

$$\frac{\Delta MP_j{}^{01}}{MP_j^{91}} = \frac{\mathrm{MP}_j^{01} - \mathrm{MP}_j^{91}}{\mathrm{MP}_j^{91}} \tag{9}$$

To simplify further, let us consider only the variation (the numerator in Equation (9)). That is,

$$\Delta MP_j{}^{01} = \sum_k \frac{POP_k^{01}}{d_{jk}^{01}} - \sum_k \frac{POP_k^{91}}{d_{jk}^{91}} \tag{10}$$

Define $\Delta\overline{MP}_j$ and $\Delta\widetilde{MP}_j$ as :

$$\Delta\overline{MP}_j{}^{01} = \sum_k \frac{POP_k^{91}}{d_{jk}^{01}} - \sum_k \frac{POP_k^{91}}{d_{jk}^{91}}$$

$$\Delta\widetilde{MP}_j{}^{01} = \sum_k \frac{POP_k^{01}}{d_{jk}^{91}} - \sum_k \frac{POP_k^{91}}{d_{jk}^{91}} \tag{11}$$

After adding and subtracting $POP_k^{91}$ in Equation (10), we obtain:

$$\Delta MP_j{}^{01} = \sum_k \frac{POP_k^{91}}{d_{jk}^{01}} - \sum_k \frac{POP_k^{91}}{d_{jk}^{91}} + \sum_k \frac{\Delta POP_k^{01}}{d_{jk}^{01}} \tag{12}$$

Rearranging Equation (12), we have:

$$\Delta MP_j{}^{01} = \sum_k \left( \frac{POP^{91}{}_k}{d_{jk}^{01}} - \frac{POP_k^{91}}{d_{jk}^{91}} \right) + \sum_k \frac{\Delta POP_k^{01}}{d_{jk}^{01}}$$
$$\leftrightarrow \tag{13}$$
$$\Delta MP_j{}^{01} = \Delta \overline{MP}_j{}^{01} + \sum_k \frac{\Delta POP_k^{01}}{d_{jk}^{01}}$$

Adding and subtracting $\left( \frac{\Delta POP_1^{01}}{d_{j1}^{91}} + \cdots + \frac{\Delta POP_{275}^{01}}{d_{j275}^{91}} \right)$ in Equation (13), give us the following:

$$\Delta MP_j{}^{01} = \Delta \overline{MP}_j{}^{01} + \Delta \widetilde{MP}_j{}^{01} + \Delta \epsilon_j{}^{01}$$

where,

$$\Delta \epsilon_j{}^{01} = \sum_k \left( \frac{\Delta POP_k^{01}}{d_{jk}^{01}} - \frac{\Delta POP_k^{01}}{d_{jk}^{91}} \right) \tag{14}$$

Finally, we multiply and divide the term $\Delta \epsilon_j$ in Equation (14) by $d_{jk}^t$, where $t = \{1991, 2001\}$, and then we rearrange it to get:

$$\Delta \epsilon_j{}^{01} = \sum_k \left( \frac{d_{jk}^{91} \Delta POP_k^{01}}{d_{jk}^{91} d_{jk}^{01}} - \frac{d_{jk}^{01} \Delta POP_k^{01}}{d_{jk}^{01} d_{jk}^{91}} \right)$$
$$\leftrightarrow \tag{15}$$
$$\Delta \epsilon_j{}^{01} = \sum_k \frac{\Delta POP_k^{01} (d_{jk}^{91} - d_{jk}^{01})}{d_{jk}^{01} d_{jk}^{91}} \ \blacksquare$$

where the residual term, $\Delta \epsilon_j$, is the analog of the so-called cross-term effect in the shift-share analysis. Having shown that, we can now take the average to get the result presented in Table 2:

$$\frac{\sum_i \frac{\Delta MP_i{}^{01}}{MP_i^{91}}}{N} = \frac{\sum_i \frac{\Delta \overline{MP}_i{}^{01}}{MP_i^{91}}}{N} + \frac{\sum_i \frac{\Delta \widetilde{MP}_i{}^{01}}{MP_i^{91}}}{N} + \frac{\sum_i \frac{\Delta \epsilon_i{}^{01}}{MP_i^{91}}}{N} \tag{16}$$

where $i = \{1, \ldots j, \ldots, N\}$ and $N = 275$.

Computing the variation between 2011 and 2001 demands more effort in terms of algebra, but the rationale is essentially the same. We start by writing the variation between 2011 and 2001 as:

$$\frac{\Delta MP_j{}^{11}}{MP_j^{01}} = \frac{\mathrm{MP}_j^{11} - \mathrm{MP}_j^{01}}{\mathrm{MP}_j^{01}}$$

Again, let us consider the numerator only as below,

$$\Delta MP_j{}^{11} = \sum_j \frac{POP_k^{11}}{d_{jk}^{11}} - \frac{POP_k^{01}}{d_{jk}^{01}} \tag{17}$$

Similarly, for $\Delta\overline{MP}_j$ and $\Delta\widetilde{MP}_j$:

$$\begin{aligned} \Delta\overline{MP}_j{}^{11} &= \sum_j \frac{POP_k^{91}}{d_{jk}^{11}} - \frac{POP_k^{91}}{d_{jk}^{01}} \\ \Delta\widetilde{MP}_j{}^{11} &= \sum_j \frac{POP_k^{11}}{d_{jk}^{91}} - \frac{POP_k^{01}}{d_{jk}^{91}} \end{aligned} \tag{18}$$

Without loss of generality and to save space, let set $k = \{1\}$. Then, we add and subtract both $\Delta\overline{MP}_j{}^{11}$and $\Delta\widetilde{MP}_j{}^{11}$ to $\Delta MP_j{}^{11}$ in Equation (17). In this context, we have:

$$\begin{aligned} \Delta MP_{j1}{}^{11} = \Delta\overline{MP}_{j1}{}^{11} + \Delta\widetilde{MP}_{j1}{}^{11} \\ + \left[\left(\frac{POP_1^{11}}{d_{j1}^{11}} - \frac{POP_1^{01}}{d_{j1}^{01}}\right) - \left(\frac{POP_1^{11}}{d_{j1}^{91}} - \frac{POP_1^{01}}{d_{j1}^{91}}\right) \right. \\ \left. + \left(\frac{POP_1^{91}}{d_{j1}^{11}} - \frac{POP_1^{91}}{d_{j1}^{01}}\right)\right] \end{aligned} \tag{19}$$

Rearranging the last term of the right-handed side of Equation (19) we finally get,

$$\Delta MP_{j1}{}^{11} = \Delta\overline{MP}_{j1}{}^{11} + \Delta\widetilde{MP}_{j1}{}^{11} + \Delta\varepsilon_{j1}{}^{11}$$

where,

$$\Delta\epsilon_{j1}{}^{11} = \left(\frac{POP_1^{11} - POP_1^{91}}{d_{j1}^{11}}\right) - \left[\left(\frac{\Delta POP_1^{11}}{d_{j1}^{91}}\right) + \left(\frac{\Delta POP_1^{01}}{d_{j1}^{01}}\right)\right] \blacksquare \tag{20}$$

**Figure B1. Market potential vs. Market potential (Population of 1991 fixed – *road effect*)**

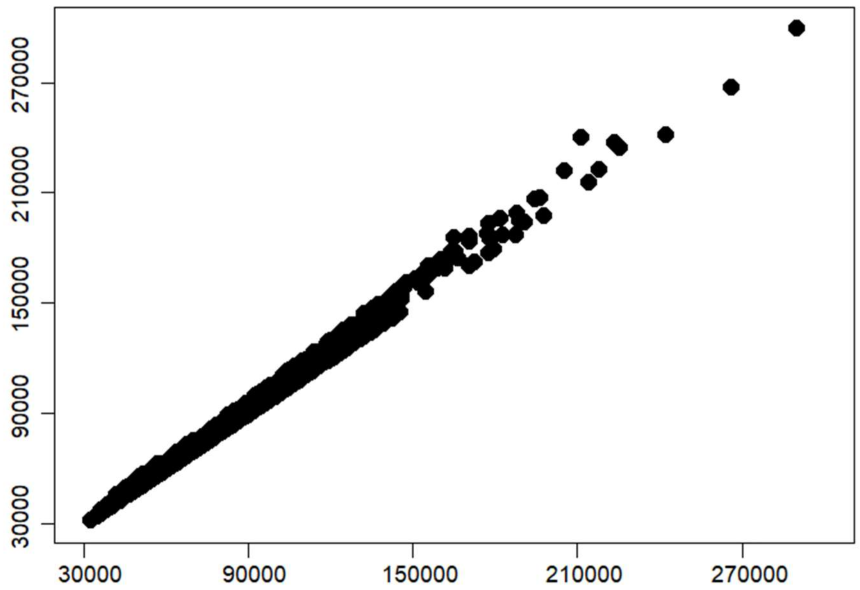
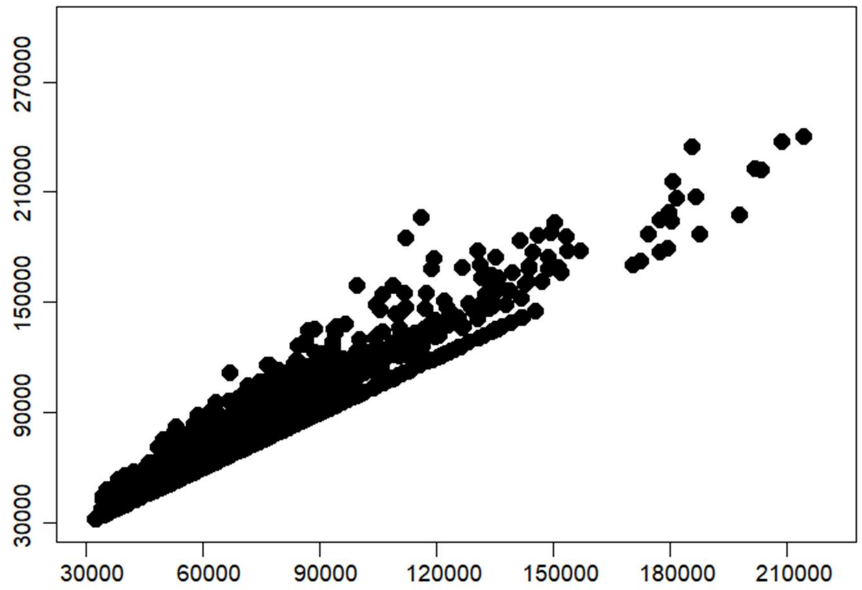



**Figure B2. Market potential vs. Market potential (Road network of 1991 fixed – *population effect*)**




**Notes:** "Market potential" is calculated as in Equation (4); "Market potential (Population of 1991 fixed – *road effect*)" refers to $\overline{MP}$ and "Market potential (Road network of 1991 fixed – *population effect*)" refers to $\widetilde{MP}$. All of these terms are defined in Section 4.3.1.

*The control function approach*

To address the endogeneity of market potential, we implement the PPML-FEIV estimator using the control function approach proposed by Lin and Wooldridge (2019). In our application, this method involves estimating the first stage – Equation 21, below – by regressing market potential on the full set of control variables described in Section 4.3.2, along with the two sets of fixed effects outlined in Section 3:

$$\begin{aligned}\log(MP_{j,t-1}) = {} & \gamma_0 + \gamma_1 CIV_{j,t-1}^{h} + \gamma_2 \log(HHI_{j,t-1}) + \gamma_3 LOC_{js,t-1} \\ & + \gamma_4 \log(WAGE_{js,t-1}) + \gamma_5 \log(POP_{j,t-1}) + \alpha_j + \alpha_{s,t} \\ & + \varepsilon_{jt-}\end{aligned} \tag{21}$$

$$h = \{1,2\}$$

where $CIV_{j,t-1}^{h}$ is one of our two composite instruments discussed in Section 3.3.2 – that are included in the equation above separately, that is, each model is exactly identified; and $\varepsilon_{jt-1}$ is the error term. Then, we include the residuals into the Poisson regression with FE specification as follows:

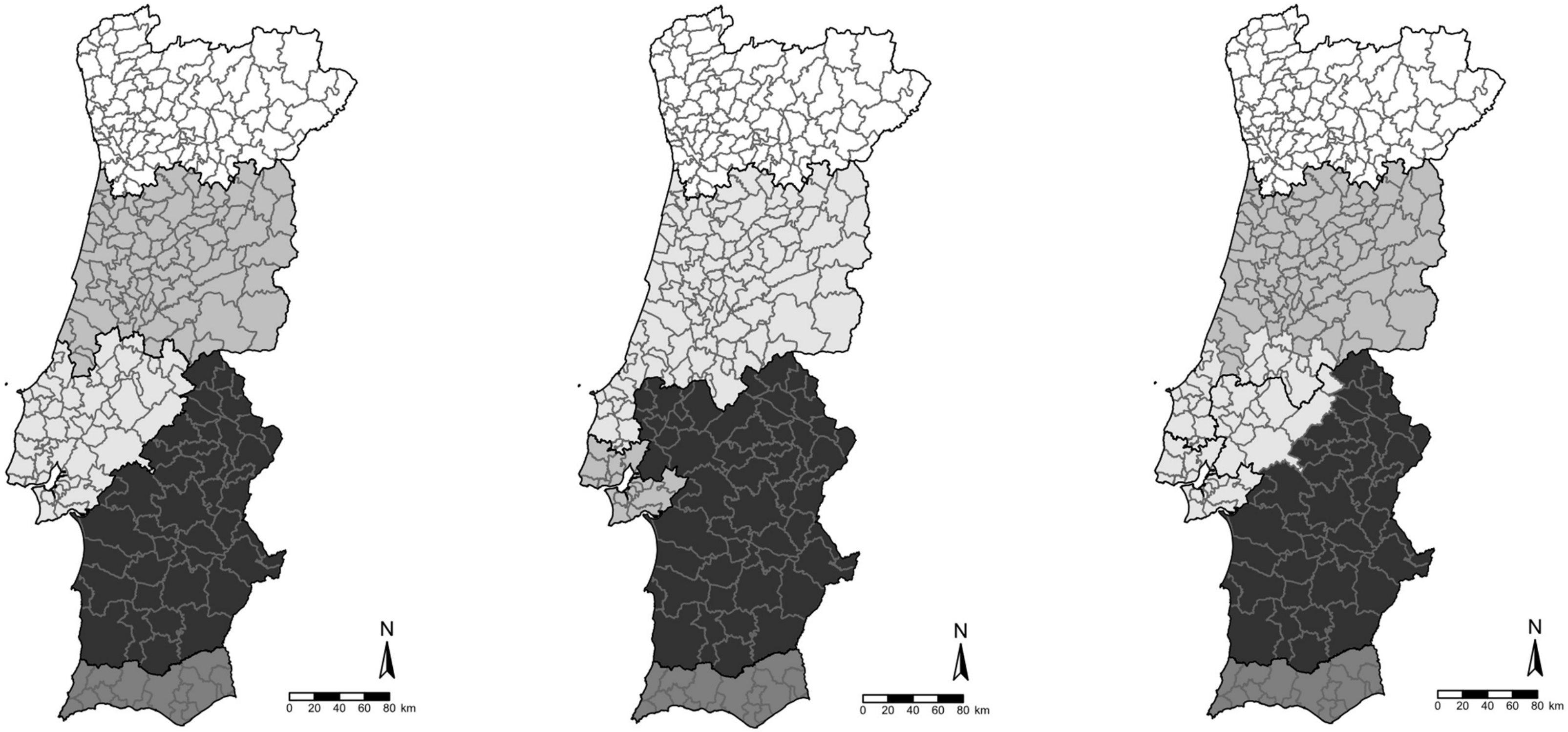
$$\begin{aligned}\mathbb{E}(n_{jst}) = {} & \exp(\beta_o + \rho\hat{\varepsilon}_{j,t} + \beta_1 \log(MP_{j,t-1}) + \beta_2 \log(HHI_{j,t-1}) \\ & + \beta_3 LOC_{js,t-1} + \beta_4 \log(WAGE_{js,t-1}) + \beta_5 \log(POP_{j,t-1}) \\ & + \beta_6 \log(HC_{j,t-1}) + \alpha_j + \alpha_{s,t})\end{aligned} \tag{22}$$

$$t = \{1992, 2002, 2012\}$$

where $\mathbb{E}(n_{jst})$ stands for the conditional expectation of $n_{jst}$; $n_{jst}$ is the number of new firms in industry $s$, location $j$ and time; $\beta_1$ is the coefficient of interest; $\hat{\varepsilon}_{j,t}$ is the predicted residual from the first stage; $\rho$ and $\beta_j$, $j = 0,1,\ldots,6$ are unknown coefficients to be estimated. To account for variation in the first stage of our standard error calculation, we use a cluster-robust bootstrap with 300 repetitions.

## Appendix C

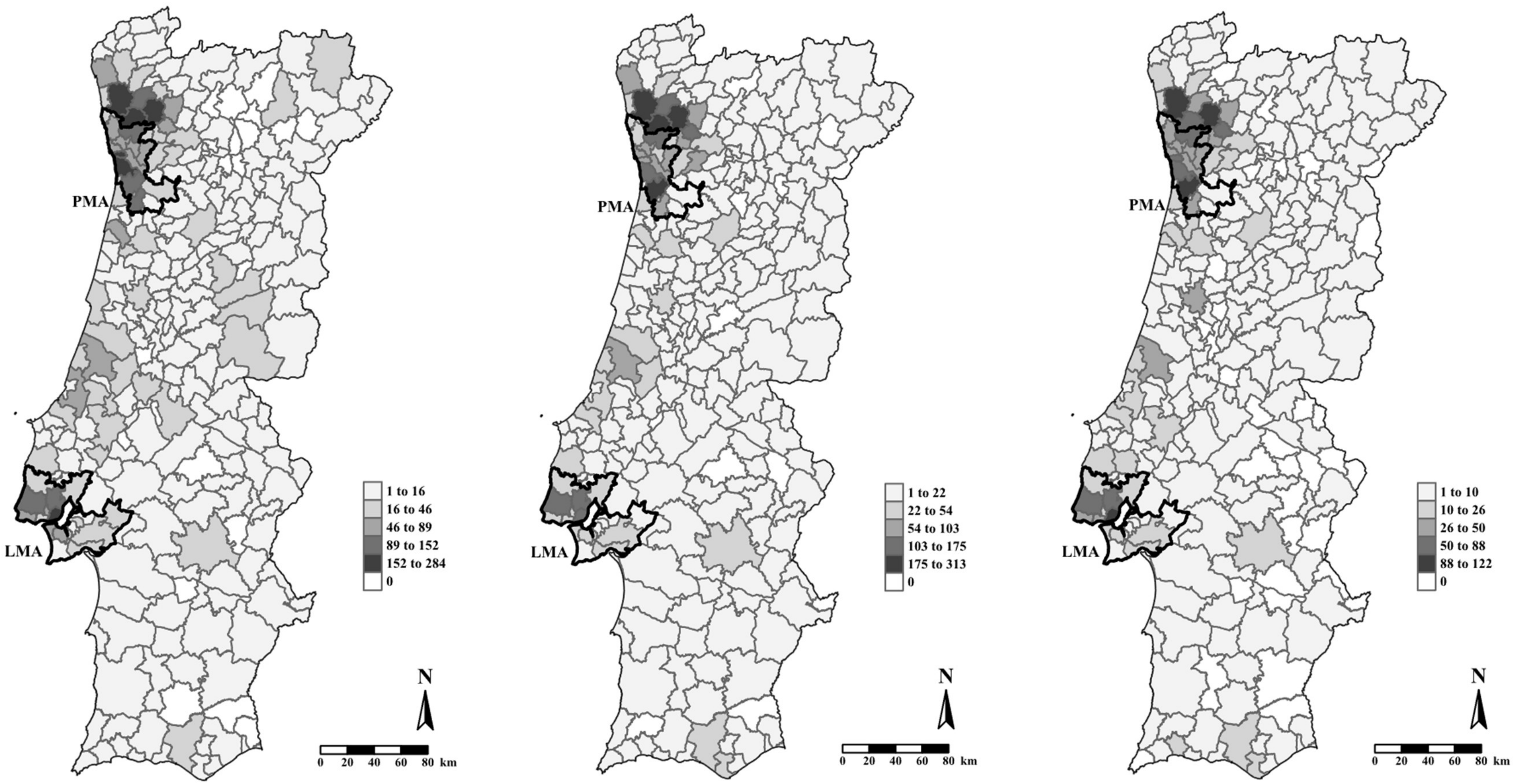
**Figure C1. NUTS2 regions in Portugal: a comparison of 1986 and 2013 classifications**




**Panel A. NUTS2 regions in 1986.** **Panel B. NUTS2 regions since 2013.** **Panel C. Comparison between NUTS2 regions.**

**Notes**. Panel A and B: The colored zones represent the municipalities within the NUTS2 region, and the bold line delineates the borders of the entire NUTS2 region. Panel C: The colored zones represent the municipalities within the NUTS2 region following the 1986 classification, while the bold line delineates the borders of the entire NUTS2 region following the 2013 classification. **Source:** Authors' calculations.

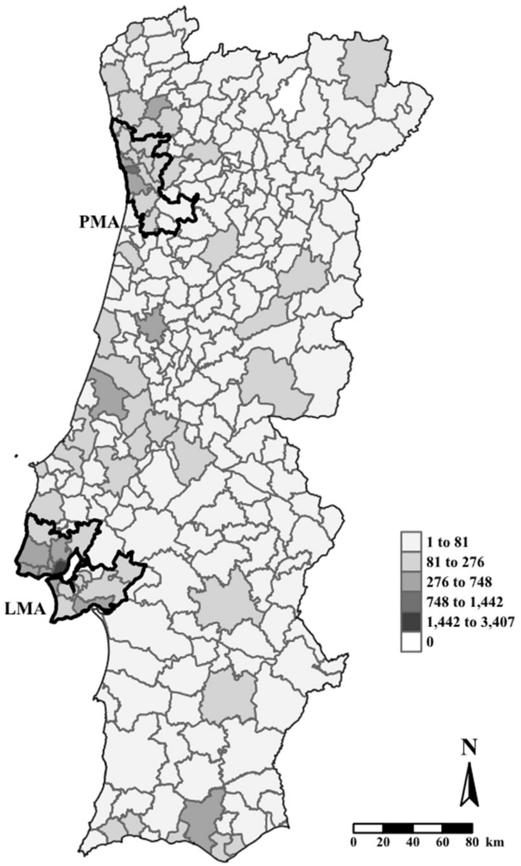
**Figure C2. Spatial distribution of manufacturing firm births by municipality**



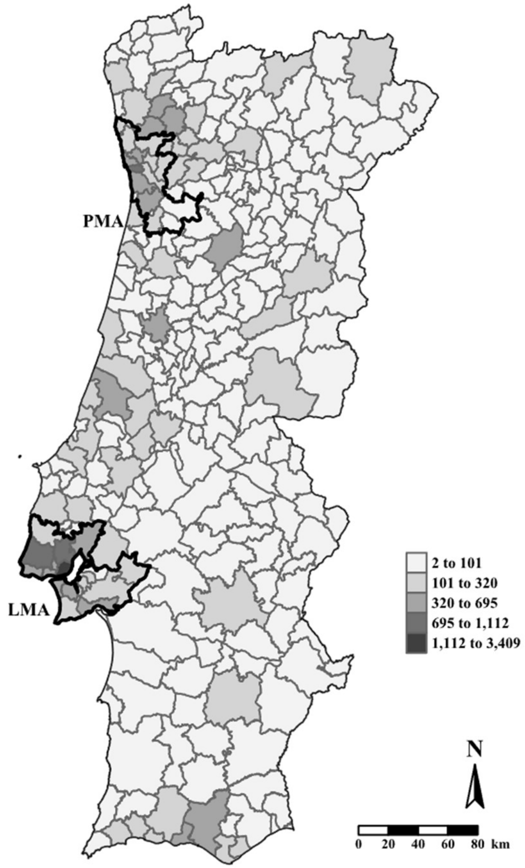

**Panel A. Number of births in 1992.** **Panel B. Number of births in 2002.** **Panel C. Number of births in 2012.**

**Notes**. "PMA" stands for Porto Metropolitan Area; "LMA" stands for Lisbon Metropolitan Area; the LMA and PMA are the only two metropolitan areas in Portugal. Maps were generated based on the "Jenks classification" in R, which identifies logical breakpoints by grouping similar values that minimize differences between values in the same cluster and maximize the differences between clusters. **Source:** Authors' calculations.

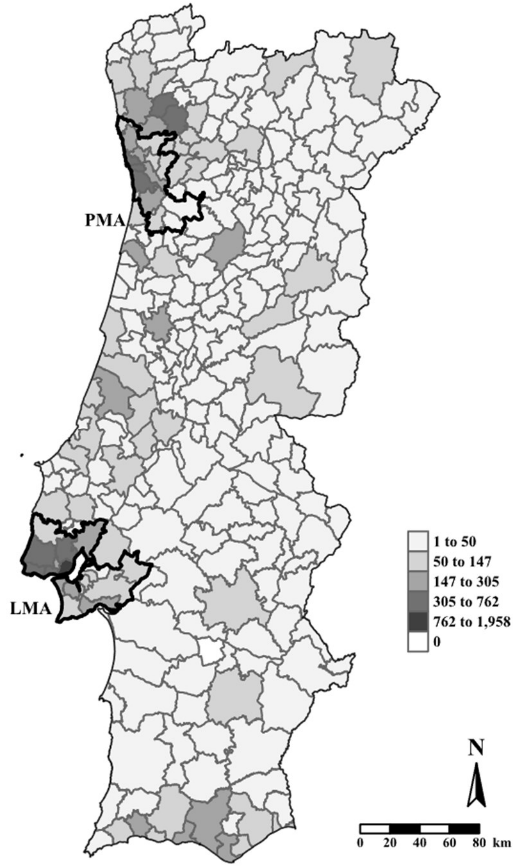
## Figure C3. Spatial distribution of services firm firths by municipality




**Panel A. Number of births in 1992.**




**Panel B. Number of births in 2002.**




**Panel C. Number of births in 2012.**

**Notes**. "PMA" stands for Porto Metropolitan Area; "LMA" stands for Lisbon Metropolitan Area; the LMA and PMA are the only two metropolitan areas in Portugal. Maps were generated based on the "Jenks classification" in R, which identifies logical breakpoints by grouping similar values that minimize differences between values in the same cluster and maximize the differences between clusters. **Source:** Authors' calculations.

## Table C1. Variables description, descriptive statistics, and previous studies

| Location determinant | Description | Source | Mean (s.d.) | Min. – Max. | Previous studies studying the effect of this location-determinant |
|---|---|---|---|---|---|
| Market potential | Distance discounted sum of the neighbor's population, where distance is computed by road travel time (minutes) | Afonso et al. (2025) | 94,247 (38,446) | 32,285 – 299,892 | Holl (2004a, 2004b, 2004c); Holl & Mariotti (2018b); Melo et al. (2010); Otsuka (2008). |
| Urbanization economies | The Herfindahl-Hirschman index | *Quadros de Pessoal* | 0.162 (0.075) | 0.04 – 0.78 | Alañón-Pardo et al. (2018); Bhat et al. (2014); Daunfeldt et al. (2013); Holl (2004a, 2004b, 2004c); Liviano & Arauzo-Carod (2013, 2014); Manjón-Antolín & Arauzo-Carod (2011). |
| Localization economies | Index of relative specialization | *Quadros de Pessoal* | 0.844 (3.583) | 0.00 – 344.37 | Alañón-Pardo et al. (2018); Alañón-Pardo & Arauzo-Carod (2013); Bhat et al. (2014); Capozza et al. (2018); Guimarães et al. (2004); Holl (2004a, 2004b, 2004c); Melo et al. (2010). |
| Market size | Total resident population | INE | 35,507 (59,466) | 1,834 – 663,394 | Holl (2004a, 2004b, 2004c); Kim et al. (2018); Holl & Mariotti (2018b); Guimarães et al. (2000); Melo et al. (2010); Manjón-Antolín & Arauzo-Carod (2011); Arauzo-Carod & Manjón-Antolín (2012); Daunfeldt et al. (2013); Mota & Brandão (2013); Bhat et al. (2014); Capozza et al. (2018) |
| Labor cost | Average real wage for each municipality and sector of economic activity (CAE) | *Quadros de Pessoal* | 633.95 (314.74) | 5.94 – 22,285 | Guimarães et al. (2000,2004); Holl (2004a, 2004b, 2004c); Otsuka (2008); Melo et al. (2010); Manjón-Antolín & Arauzo-Carod (2011); Mota & Brandão (2013); Kim et al. (2018). |
| Human capital | Percentage of the resident population with complete higher education | INE | 5.72 (4.78) | 0.38 – 33.63 | Alañón-Pardo et al. (2018); Alañón-Pardo & Arauzo-Carod (2013); Arauzo-Carod & Manjón-Antolín (2012); Daunfeldt et al. (2013); Guimarães et al. (2000); Holl (2004a, 2004b, 2004c); Manjón-Antolín & Arauzo-Carod (2011); Mota & Brandão (2013). |

Source: Authors' compilation.

## Table C2. Correlation Matrix

| Variables | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) | (13) | (14) | (15) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (1) Log of market potential (complete measure) | 1.000 | | | | | | | | | | | | | | |
| (2) Log of market potential ($\overline{MP}$) | 0.997 | 1.000 | | | | | | | | | | | | | |
| (3) Log of market potential ($\widetilde{MP}$) | 0.953 | 0.968 | 1.000 | | | | | | | | | | | | |
| (4) Log of market potential ($\overline{MP}$) (exc. metro area) | 0.813 | 0.809 | 0.764 | 1.000 | | | | | | | | | | | |
| (5) Log of urbanization economies | -0.354 | -0.338 | -0.294 | -0.277 | 1.000 | | | | | | | | | | |
| (6) Localization economies | 0.012 | 0.012 | 0.014 | -0.005 | -0.028 | 1.000 | | | | | | | | | |
| (7) Log of labor cost | 0.257 | 0.250 | 0.227 | 0.135 | -0.159 | 0.091 | 1.000 | | | | | | | | |
| (8) Log of market size | 0.564 | 0.571 | 0.590 | 0.323 | -0.294 | 0.012 | 0.292 | 1.000 | | | | | | | |
| (9) Log of human capital | 0.265 | 0.261 | 0.238 | 0.137 | -0.222 | 0.009 | 0.536 | 0.385 | 1.000 | | | | | | |
| (10) CIV1: the fraction of MW+EW – The Roman roads (straight-line distance to the nearest road) | 0.476 | 0.492 | 0.478 | 0.304 | -0.188 | 0.010 | 0.086 | 0.184 | 0.080 | 1.000 | | | | | |
| (11) CIV2: the fraction of MW+EW – The 18th-century itineraries(straight-line distance to the nearest road) | 0.394 | 0.414 | 0.438 | 0.445 | -0.207 | 0.002 | 0.040 | 0.082 | 0.032 | 0.713 | 1.000 | | | | |
| (12) CIV1a: the fraction of MW+EW – The Roman roads (excluding neighbors of neighbors) | 0.357 | 0.372 | 0.350 | 0.249 | -0.132 | -0.001 | 0.047 | 0.094 | 0.047 | 0.950 | 0.711 | 1.000 | | | |
| (13) CIV2a: the fraction of MW+EW – The 18th-century itineraries (excluding neighbors of neighbors) | 0.289 | 0.308 | 0.324 | 0.374 | -0.139 | -0.001 | 0.010 | 0.003 | -0.008 | 0.632 | 0.953 | 0.685 | 1.000 | | |
| (14) CIV1b: the fraction of MW+EW – The Roman roads (length of the network) | 0.317 | 0.319 | 0.247 | 0.351 | -0.016 | -0.005 | -0.004 | 0.133 | -0.006 | 0.541 | 0.324 | 0.582 | 0.331 | 1.000 | |
| (15) CIV2b: the fraction of MW+EW – The 18th-century itineraries (length of the network) | 0.240 | 0.256 | 0.278 | 0.464 | -0.118 | -0.008 | -0.033 | -0.062 | -0.043 | 0.301 | 0.829 | 0.362 | 0.834 | 0.239 | 1.000 |

**Notes:** "MW+EW" stands for motorways and expressways; "exc. metro area" stands for excluding municipalities from the two metropolitan areas of Lisbon and Porto.

## Table C3. PPML-FE estimation: Reduced-form regression

| Dependent variables: Count of firm births | (1) 1-year interval | (2) 5-year interval |
|---|---|---|
| CIV1: the fraction of MW+EW | 17.566*** | 11.934*** |
| *The Roman major roads* | (5.810) | (4.018) |
| CIV2: the fraction of MW+EW | 1.683*** | 1.448*** |
| *The 18th-century itineraries* | (0.649) | (0.417) |
| Observations | 24,475 | 25,850 |

**Notes.** * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. In parentheses: cluster-robust standard errors. "1-year interval" stands for the count of establishment births measured in 1992, 2002, and 2012. "5-year interval" stands for the sum of establishment births measured in 1992-1996, 2002-2006, and 2012-2016. All regressions include control variables (urbanization and localization economies, labor cost, market size, and human capital), municipality fixed effects, and Sector of Activity (CAE) × Year fixed effects. "MW+EW" stands for motorways and expressways. Control variables are included in logarithmic form, except for localization economies, since they contain a considerable number of zeros.

## Table C4. PPML-FEIV estimation: Progressively including controls

| Dependent variables: Count of firm births | (1) 1-year interval | (2) 5-year interval | (3) 1-year interval | (4) 5-year interval | (5) 1-year interval | (6) 5-year interval |
|---|---|---|---|---|---|---|
| Market potential (*road effect*) | | | | | | |
| CIV1: the fraction of MW+EW | 1.371 | 1.151** | 1.553* | 0.914* | 1.878*** | 1.228** |
| *The Roman major roads* | (0.845) | (0.470) | (0.882) | (0.490) | (0.691) | (0.480) |
| CIV2: the fraction of MW+EW | 1.594* | 1.459*** | 1.536* | 1.183** | 1.597** | 1.339*** |
| *The 18th-century itineraries* | (0.828) | (0.486) | (0.865) | (0.517) | (0.662) | (0.441) |
| *Set of Fixed effects* | Yes | Yes | Yes | Yes | Yes | Yes |
| *Controls* | | | | | | |
| *Agglomeration economies* | No | No | Yes | Yes | Yes | Yes |
| *Labour market* | No | No | No | No | Yes | Yes |
| | First-stage results: market potential (*road effect*) as dependent variable | | | | | |
| CIV1 | 7.334*** | | 7.330*** | | 7.119*** | |
| | (0.970) | | (0.971) | | (0.953) | |
| KP F-statistic | 57.23 | | 57.02 | | 55.80 | |
| CIV1.1 | 9.553*** | | 9.544*** | | 9.296*** | |
| | (1.201) | | (1.203) | | (1.179) | |
| KP F-statistic | 63.26 | | 62.91 | | 62.18 | |
| CIV2 | 1.062*** | | 1.062*** | | 1.056*** | |
| | (0.121) | | (0.122) | | (0.118) | |
| KP F-statistic | 76.47 | | 76.13 | | 79.78 | |
| Observations | 24,475 | 25,850 | 24,475 | 25,850 | 24,475 | 25,850 |

**Notes.** * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. In parentheses: Bootstrap (cluster) standard errors (300 reps.) and cluster-robust for first-stage coefficient. "1-year interval" stands for the count of establishment births measured in 1992, 2002, and 2012. "5-year interval" stands for the sum of establishment births measured in 1992-1996, 2002-2006, and 2012-2016. The set of Fixed effects includes municipality fixed effects and Sector of Activity (CAE) × Year fixed effects. Agglomeration economies include urbanization and localization economies. Labour market includes labor cost, market size, and human capital. "MW+EW" stands for motorways and expressways. Control variables are included in logarithmic form, except for localization economies, since they contain a considerable number of zeros.

## Table C5. PPML-FEIV: Market potential (complete measure)

| Dependent variables: Count of firm births | (1) 1-year interval | (2) 5-year interval |
|---|---|---|
| Market potential | | |
| CIV1: the fraction of MW+EW | 1.852*** | 1.216*** |
| *The Roman major roads* | (0.683) | (0.457) |
| CIV2: the fraction of MW+EW | 1.517** | 1.263*** |
| *The 18th-century itineraries* | (0.633) | (0.414) |
| | First-stage results: Market potential as the dependent variable | |
| CIV1 | 9.454*** | |
| | (1.180) | |
| KP F-statistic | 64.22 | |
| CIV2 | 1.107*** | |
| | (0.117) | |
| KP F-statistic | 89.12 | |
| Observations | 21,054 | 22,264 |

**Notes.** * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. In parentheses: Bootstrap (cluster) standard errors (300 reps.) and cluster-robust for first-stage coefficient. "1-year interval" stands for the count of establishment births measured in 1992, 2002, and 2012. "5-year interval" stands for the sum of establishment births measured in 1992-1996, 2002-2006, and 2012-2016. All regressions include control variables (urbanization and localization economies, labor cost, market size, and human capital), municipality fixed effects, and Sector of Activity (CAE) × Year fixed effects. "MW+EW" stands for motorways and expressways. "KP F-statistic" stands for the Kleibergen-Paap (2006) rk Wald F statistic. Market potential and control variables are included in logarithmic form, except for localization economies, since they contain a considerable number of zeros.

.

## Table C6. PPML-FEIV: Excluding municipalities from LMA and PMA

| Dependent variables: Count of firm births | (1) 1-year interval | (2) 5-year interval |
|---|---|---|
| Market potential (*road effect*) [a] | | |
| CIV1: the fraction of MW+EW | 3.513** | 2.928*** |
| *The Roman major roads* | (1.458) | (1.029) |
| CIV2: the fraction of MW+EW | 3.753** | 3.598*** |
| *The 18th-century itineraries* | (1.533) | (1.202) |
| | First-stage results: Market potential (*road effect*) [a] as the dependent variable | |
| CIV1 | 5.620*** | |
| | (1.153) | |
| KP F-statistic | 23.76 | |
| CIV2 | 0.579*** | |
| | (0.120) | |
| KP F-statistic | 23.29 | |
| Observations | 15,575 | 16,450 |

**Notes.** * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. In parentheses: Bootstrap (cluster) standard errors (300 reps.) and cluster-robust for first-stage coefficient. "1-year interval" stands for the count of establishment births measured in 1992, 2002, and 2012. "5-year interval" stands for the sum of establishment births measured in 1992-1996, 2002-2006, and 2012-2016. [a] The computation of market potential excludes municipalities located within the metropolitan areas of Lisbon and Porto as potential destinations. All regressions include control variables (urbanization and localization economies, labor cost, market size, and human capital), municipality fixed effects, and Sector of Activity (CAE) × Year fixed effects. "MW+EW" stands for motorways and expressways. "KP F-statistic" stands for the Kleibergen-Paap (2006) rk Wald F statistic. Market potential and control variables are included in logarithmic form, except for localization economies, since they contain a considerable number of zeros.

## Table C7. PPML-FEIV: Manufacturing vs. Services

| Dependent variables: Count of firm births | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | 1-year Interval | 5-year interval | 1-year interval | 5-year interval |
| | Manufacturing | | Services | |
| Market potential (*road effect*) | | | | |
| CIV1: the fraction of MW+EW | -0.150 | -0.187 | 2.508*** | 1.607*** |
| *The Roman major roads* | (1.083) | (0.845) | (0.847) | (0.478) |
| CIV2: the fraction of MW+EW | -0.104 | -0.331 | 1.694** | 1.373*** |
| *The 18th-century itineraries* | (1.091) | (0.776) | (0.759) | (0.419) |
| | First-stage results: Market potential (*road effect*) as the dependent variable | | | |
| CIV1 | 9.299*** | | 9.294*** | |
| | (1.178) | | (1.180) | |
| KP F-statistic | 62.26 | | 62.08 | |
| CIV2 | 1.056*** | | 1.056*** | |
| | (0.118) | | (0.118) | |
| KP F-statistic | 79.78 | | 79.83 | |
| Observations | 12,831 | 14,300 | 9,075 | 9,075 |

**Notes.** * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. In parentheses: Bootstrap (cluster) standard errors (300 reps.) and cluster-robust for first-stage coefficient. "1-year interval" stands for the count of establishment births measured in 1992, 2002, and 2012. "5-year interval" stands for the sum of establishment births measured in 1992-1996, 2002-2006, and 2012-2016. All regressions include control variables (urbanization and localization economies, labor cost, market size, and human capital), municipality fixed effects, and Sector of Activity (CAE) × Year fixed effects. "MW+EW" stands for motorways and expressways. "KP F-statistic" stands for the Kleibergen-Paap (2006) rk Wald F statistic. Market potential and control variables are included in logarithmic form, except for localization economies, since they contain a considerable number of zeros.

## Table C8. PPML-FEIV: Including local geography × Year fixed effects

| Dependent variables: Count of firm births | (1) 1-year interval | (2) 5-year interval |
|---|---|---|
| Market potential (*road effect*) | | |
| CIV1: the fraction of MW+EW | 1.756** | 1.341** |
| *The Roman major roads* | (0.787) | (0.546) |
| CIV2: the fraction of MW+EW | 1.594** | 1.511*** |
| *The 18th-century itineraries* | (0.750) | (0.508) |
| | First-stage results: Market potential (*road effect*) as the dependent variable | |
| CIV1 | 8.545*** | |
| | (1.214) | |
| KP F-statistic | 49.58 | |
| CIV2 | 0.986*** | |
| | (0.110) | |
| KP F-statistic | 68.10 | |
| Observations | 24,475 | 25,850 |

**Notes.** * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. In parentheses: Bootstrap (cluster) standard errors (300 reps.) and cluster-robust for first-stage coefficient. "1-year interval" stands for the count of establishment births measured in 1992, 2002, and 2012. "5-year interval" stands for the sum of establishment births measured in 1992-1996, 2002-2006, and 2012-2016. All regressions include control variables (urbanization and localization economies, labor cost, market size, and human capital), municipality fixed effects, local geography × Year fixed effects, and Sector of Activity (CAE) × Year fixed effects. Local geography is defined as the standard deviation of terrain elevation. "MW+EW" stands for motorways and expressways. "KP F-statistic" stands for the Kleibergen-Paap (2006) rk Wald F statistic. Market potential and control variables are included in logarithmic form, except for localization economies, since they contain a considerable number of zeros.

## Table C9. PPML-FEIV: Extending to 8-year interval

| Dependent variables: Count of firm births | (1) 1-year interval | (2) 8-year interval |
|---|---|---|
| Market potential (*road effect*) | | |
| CIV1: the fraction of MW+EW | 1.878*** | 0.863 |
| *The Roman major roads* | (0.691) | (0.527) |
| | | |
| CIV2: the fraction of MW+EW | 1.597** | 1.229*** |
| *The 18th-century itineraries* | (0.662) | (0.446) |
| | First-stage results: Market potential (*road effect*) as the dependent variable | |
| CIV1 | 9.296*** | |
| | (1.179) | |
| KP F-statistic | 62.18 | |
| | | |
| CIV2 | 1.056*** | |
| | (0.118) | |
| KP F-statistic | 79.78 | |
| Observations | 24,475 | 25,850 |

**Notes.** * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. In parentheses: Bootstrap (cluster) standard errors (300 reps.) and cluster-robust for first-stage coefficient. "1-year interval" stands for the count of establishment births measured in 1992, 2002, and 2012. "8-year interval" stands for the sum of establishment births measured in 1992-1999, 2002-2009, and 2012-2019. All regressions include control variables (urbanization and localization economies, labor cost, market size, and human capital), municipality fixed effects, and Sector of Activity (CAE) × Year fixed effects."MW+EW" stands for motorways and expressways. "KP F-statistic" stands for the Kleibergen-Paap (2006) rk Wald F statistic. Market potential and control variables are included in logarithmic form, except for localization economies, since they contain a considerable number of zeros.

.

## Table C10. PPML-FEIV: Dummies for road accessibility

| Dependent variables: Count of firm births | (1) 1-year interval | (2) 5-year interval |
|---|---|---|
| Market potential | | |
| CIV1: the fraction of MW+EW | 1.851*** | 1.216*** |
| *The Roman major roads* | (0.657) | (0.460) |
| | | |
| CIV2: the fraction of MW+EW | 1.570** | 1.310*** |
| *The 18th-century itineraries* | (0.650) | (0.429) |
| | First-stage results: Market potential as the dependent variable | |
| CIV1 | 9.767*** | |
| | (1.070) | |
| KP F-statistic | 83.34 | |
| | | |
| CIV2 | 1.097*** | |
| | (0.110) | |
| KP F-statistic | 99.15 | |
| Observations | 21,054 | 22,264 |

**Notes.** * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. In parentheses: Bootstrap (cluster) standard errors (300 reps.) and cluster-robust for first-stage coefficient. "1-year interval" stands for the count of establishment births measured in 1992, 2002, and 2012. "5-year interval" stands for the sum of establishment births measured in 1992-1996, 2002-2006, and 2012-2016. All regressions include control variables (urbanization and localization economies, labor cost, market size, human capital, and the dummies for road accessibility), municipality fixed effects, and Sector of Activity (CAE) × Year fixed effects. The dummies for road accessibility indicate whether a municipality falls within specific distance ranges (i.e., less than 10 km, 10–20 km, 20–50 km, and over 50 km) from the nearest motorway ramp. "MW+EW" stands for motorways and expressways. "KP F-statistic" stands for the Kleibergen-Paap (2006) rk Wald F statistic. Market potential and control variables are included in logarithmic form, except for the dummies and the localization economies, since they contain a considerable number of zeros.

## Appendix D

**Error! Reference source not found.** reports the complete PPML-FEIV estimation presented in Table 3. All explanatory variables are included in logarithmic form, interpreted as elasticities, except for localization economies since they contain a considerable number of zeros, interpreted as semi-elasticity. As mentioned in Section 5.1, we instrumented market potential individually with the two CIVs described in Section 3.2.2. What follows is a brief overview of these estimates regarding control variables. Note, however, that the results from the estimated coefficients of control variables should be interpreted cautiously because they are likely affected by endogeneity bias – as discussed in Section 3.2.

Overall, we can conclude that some of them have a similar performance for firm births throughout all specifications. More specifically, localization economies, human capital, and market size are always significant, except for labor cost, which is significant only for the 5-year interval. However, while market size has expected signs, labor cost and human capital never have expected signs – the former is supposed to be negative, while the latter is supposed to be positive. Even though a possible explanation for this behavior of human capital may be the aggregate nature of the indicator, which does not allow us to evaluate the importance of some specific skills when firms decide where to (re)locate (Mota & Brandão, 2013); it is hard to justify why labor cost is persistently positive.[1] Perhaps higher wages signal more qualified workers or a higher level of demand – the higher the wage, the higher the available income.

Concerning agglomeration economies, we observe that urbanization economies and localization economies show opposite directions. On the one hand, estimates for the Herfindahl index are negative for firm births, although significant only at 10% for the 5-year interval. A similar result was found by Holl (2004b). This might suggest that a less concentrated economic structure encourages firm birth. In more diversified areas, firms are in contact with different production processes from a variety of activities, benefiting new firms in how to learn more about their best production technology. On the other hand, own-sector specialization positively influences firm births. This result may suggest a low

[1] We estimate the same equations without using the correction for wages mentioned in Section 4.3.2 – i.e., using the "actual" average real wage, which includes zero values for certain pairs of sectors and municipalities. Results are largely similar, and we still find unexpected positive signs regarding the coefficient for labor costs (available upon request).

level of competition within the same economic activities, which increases the possibility of new entrants. Estimates for localization economies consistently show statistical significance across all regressions. If we look at columns 1 to 3 and 2 to 4, for instance, a 10% increase in urbanization economies leads to approximately a 0.8% decrease in the number of births for the 1-year interval and a 0.7% decrease for the 5-year interval. For localization economies, the estimated coefficient means that increasing the specialization index by 1 unit increases the number of births for 1-year and 5-year intervals by 19.0% and 14.7%, respectively.[2]

In general, these results are in line with previous research. When used as a proxy for urbanization, the Herfindahl index usually shows a negative impact on location choice (Bhat et al., 2014; Holl 2004a, 2004b; Liviano & Arauzo-Carod, 2013), while for localization economies, the specialization index presents a positive impact (Alañón-Pardo & Arauzo-Carod, 2013; Bhat et al., 2014; Holl, 2004a; Melo et al., 2010).

**Table D1. PPML-FEIV: Results for control variables**

| Dependent variables: Count of firm births | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | 1-year interval | 5-year interval | 1-year interval | 5-year interval |
| | CIV1: the fraction of MW+EW The Roman major roads | | CIV2: the fraction of MW+EW The 18th-century itineraries | |
| Urbanization economies | -0.081* | -0.067* | -0.086* | -0.066* |
| | (0.049) | (0.036) | (0.048) | (0.036) |
| Localization economies | 0.174*** | 0.137*** | 0.174*** | 0.137*** |
| | (0.011) | (0.017) | (0.011) | (0.017) |
| Labor cost | 0.228* | 0.273*** | 0.229* | 0.273*** |
| | (0.123) | (0.086) | (0.123) | (0.086) |
| Market size | 0.816*** | 0.412*** | 0.814*** | 0.426*** |
| | (0.114) | (0.081) | (0.115) | (0.079) |
| Human capital | -0.338*** | -0.322*** | -0.340*** | -0.321*** |
| | (0.092) | (0.074) | (0.092) | (0.074) |
| Observations | 24,475 | 25,850 | 24,475 | 25,850 |

**Notes.** * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. In parentheses: Bootstrap (cluster) standard errors (300 reps.). "1-year interval" stands for the count of establishment births measured in 1992, 2002, and 2012. "5-year interval" stands for the sum of establishment births measured in 1992-1996, 2002-2006, and 2012-2016. All regression includes market potential, the municipality fixed effects, and Sector of Activity (CAE) × Year fixed effects. "MW+EW"

[2] That is $100 * \exp(0.174) - 1)\%$ and $100 * \exp(0.137) - 1)\%$.

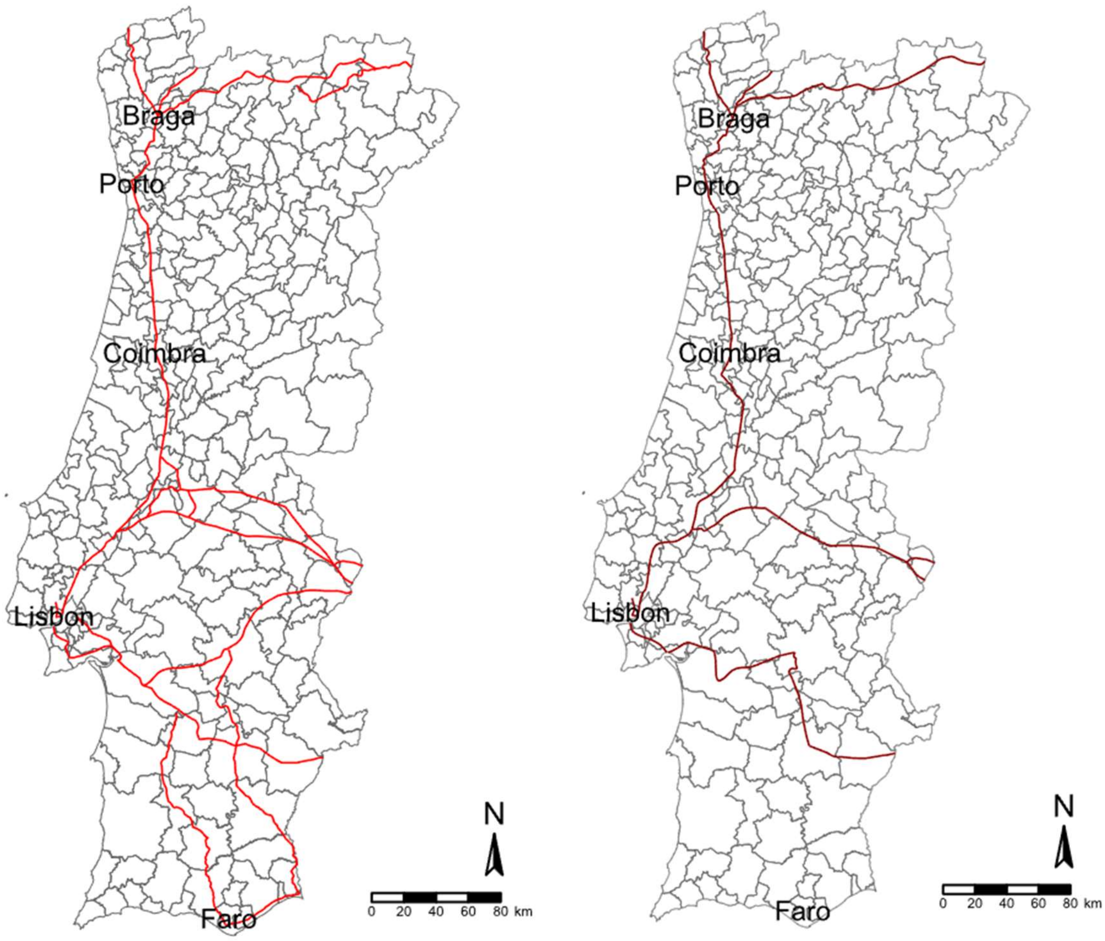
stands for motorways and expressways. Market potential and control variables are included in logarithmic form except for localization economies since they contain a considerable number of zeros.

**Appendix E**

As noted in the text, the *Mercator-e* and *Itiner-e* projects (Brughmans et al., 2024) provide a recent, comprehensive digital atlas of ancient and historical roads that have not yet been widely used in the literature as instruments for modern transport infrastructure, in contrast to the extensive use of data from McCormick et al. (2013). We present a comparison between the two datasets, with particular attention to the segments corresponding to present-day Portugal. Focusing on Roman major roads, McCormick et al. (2013) report a total length of 1,085 km for the Roman road network in Portugal. In contrast, Brughmans et al. (2024) identify a total length of approximately 1,875 km for the Roman major roads crossing Portugal. The primary differences between these datasets are concentrated in the NUTS2 regions of Alentejo and Algarve – essentially the southern part of the country, roughly between Lisbon and Faro – as illustrated in Figure E1 below. Regarding accessibility, the average distance of a municipality centroid to the nearest Roman major roads is 31.7 km when using data from McCormick et al. (2013) and 24.3 km when using data from Brughmans et al. (2024). The correlation between these two networks is approximately 0.74.

Table E1 presents the PPML-FEIV estimations, analogous to those in Tables 3, 4, and 5, but now using data on Roman major roads from Brughmans et al. (2024). Specifically, CIV1.1 corresponds to CIV1, CIV1a.1 corresponds to CIV1a, and CIV1b.1 corresponds to CIV1b. Overall, the results are very similar, although they tend to be slightly lower in magnitude and estimated with less precision than those based on McCormick et al. (2013). Out of the six-point estimates presented in Table E1, four are statistically significant at the 5% level and one at the 10% level. The F-statistics remain well above the conventional threshold, ranging from 37.6 to 61.0. In sum, these findings suggest that using alternative data for historical Roman roads does not significantly influence our results. The overall effect of market potential on firm births remains consistent, with a 10% increase in market potential leading to an increase in firm births ranging from approximately 11% to 18%. This range is particularly close to our baseline results presented in Section 5.2 and the robustness analysis in Section 5.2.1, which essentially ranges from 12% to 19%.

**Figure E1.The Roman major roads**




**Panel A. From Brughmans et al. (2024)** **Panel B. From McCormick et al. (2013)**

**Source**: Authors' calculations are based on data from the respective studies cited in panels A and B.

## Table E1.PPML-FEIV estimation: using alternative Roman roads

| Dependent variables: Count of firm births | (1) 1-year interval | (2) 5-year interval |
|---|---|---|
| Market potential (*road effect*) | | |
| CIV1.1: the fraction of MW+EW | 1.570** | 0.881* |
| *The Roman major roads* | (0.710) | (0.504) |
| *Brughmans et al. (2024)* | | |
| CIV1a.1: the fraction of MW+EW | 1.761** | 1.081** |
| *The Roman major roads* | (0.775) | (0.537) |
| *Brughmans et al. (2024)* | | |
| CIV1b.1: the fraction of MW+EW | 1.548** | 0.749 |
| *The Roman major roads* | (0.660) | (0.499) |
| *Brughmans et al. (2024)* | | |
| | First-stage results: Market potential (*road effect*) as the dependent variable | |
| CIV1.1 | 7.119*** | |
| | (0.953) | |
| KP F-statistic | 55.80 | |
| CIV1a.1 | 6.689*** | |
| | (1.091) | |
| KP F-statistic | 37.60 | |
| CIV1b.1 | 0.001*** | |
| | (0.0001) | |
| KP F-statistic | 61.04 | |
| Observations | 24,475 | 25,850 |

**Notes.** * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. In parentheses: Bootstrap (cluster) standard errors (300 reps.) and cluster-robust for first-stage coefficient. "1-year interval" stands for the count of establishment births measured in 1992, 2002, and 2012. "5-year interval" stands for the sum of establishment births measured in 1992-1996, 2002-2006, and 2012-2016. All regressions include control variables (urbanization and localization economies, labor cost, market size, and human capital), municipality fixed effects, and Sector of Activity (CAE) × Year fixed effects."MW+EW" stands for motorways and expressways. "KP F-statistic" stands for the Kleibergen-Paap (2006) rk Wald F statistic. Market potential and control variables are included in logarithmic form, except for localization economies, since they contain a considerable number of zeros.