Essays in Applied Economics

Marta Santagata

Department of Economics (DIEC)

University of Genoa

Supervisor

Maurizio Conti - University of Genoa

Anna Bottasso - University of Genoa

In partial fulfillment of the requirements for the degree of

PhD Program in Economics

2020

Contents

Preface

This work includes the results of a broad research carried out during the PhD course in Economics at the University of Genova.

The first part of this work consists of two articles that can be ascribed to the literature on the long and short run economic impact of infrastructure.

In particular, the first article is devoted to investigate the long-term impact of the Roman road network on today's propensity to export following a micro funded model by Duranton et al. (2014) that links roads to trade.
The paper main result is that, controlling for possible determinants of propensity to trade and for the lenght and the location of Roman road system dating back to 117 A.D., the latter still influences today's propensity to export of the Italian NUTS-2 regions.
Moreover, we investigate two potential channels underlying this result that have been highlighted in the previous literature. In particular, results in this paper confirm the well supported hypothesis in the literature that Roman roads are correlated with today's roads even though they were not established on the same route. However our paper does not lend support to the idea that Roman roads foster current trade by influencing social capital.

The second article[1] deals on the literature on the economic impact of transport infrastructure, and in particular on the role that road infrastructure can have on innovative regional capacity.
We follow the seminal contribution by Agrawal et al. (2017) and we estimate a model of "roads and innovation" where the innovative activity in 1988 is linked to the length of motorways system in 1983, in order to investigate the impact of motorways endowment on the innovative capacity in each Italian NUTS-3 region.
The main challenging issue about the estimation of our model arises from the possible endogeneity of highways stock. To deal with this problem, we follow the "historical instrumental variable" approach by using the length of the ancient Roman roads system dating back to 117 A.D. as an instrument for the length of current motorways. Overall, our Instrumental Variable estimates indicate that 1983 highways network has a positive and significant impact on 1988 innovative capacity.
Moreover, we find a declining role for highways over time. Furthermore, results suggest a spatial reorganization of economic activity rather than a pure net economic effect.

The second part of this work contains an article[2] that belongs to the literature on the multidimensional indexes of well-being.
In particular, following the growing interest in new and better measures of develop-

[1]This article is the result of joint work with Anna Bottasso, Maurizio Conti and Simone Robbiano

[2]This article has been published in *Social Indicator Research* in 2019, joint with Enrico Ivaldi and Riccardo Soliani (https://doi.org/10.1007/s11205-017-1815-x)

ment, we elaborate an index of development for the Republics that gained independence after the Soviet Union broke up.

We base our analysis on a set of variables from the World Development Indicators database released by the World Bank. We select the variables through principal component analysis and we calculate the index using factorial analysis.

Therefore, following the well supported hypothesis in the literature that good governance has a key positive influence on development, we compare this index with a proposed index of governance, elaborating data from the Worldwide Governance Indicators database of the World Bank. As expected, the correlation between our index of development and our index of governance is high.

Finally we perform a cluster analysis to group country according to the two indices.

Tempus fugit, Roman legacies do not: the long-run effect of Roman Roads on the propensity to trade

1 Introduction

The persistence of historical events has been analysed in various fields, with reference to different places and periods. The importance of this literature is underlined by the many important works on the long-term impact of different historical events, as documented by various surveys (see Nunn, 2009, 2014; Michalopulos and Papaioannou, 2017).

Researchers have focused on several events in the past, including European colonial rule, increased international trade, the outbreak of wars, involuntary population movements, religious events and innovative activities.

Moreover, great effort has been made to understand the mechanisms underlying persistence and different possible channels have been explored, e.g. multiple equilibria and path dependence, domestic institutions, cultural values and beliefs, and genetic traits (Nunn, 2014).

Looking to the past, searching for relationships between history and the present, is of great importance to the appropriate formulation of policy measures. Indeed, understanding the extent to which levels of contemporary economic development are still shaped by history is of fundamental importance to ensure the long-term sustainability of prosperity (Wahl, 2017).

This paper aims to offer a new contribution in this literature, by assessing whether the impact of historical Roman infrastructure persists today in defining the propensity to trade between regions of the same country and by investigating the possible mechanisms underlying persistence.

More specifically, this work is related to the more recent strand of literature on the very long-term impact of historical infrastructure on recent outcomes. Indeed, this latter strand can be seen as a development of literature on the persistence of history, and the Roman world and its heritage find a great deal of space within it. This is especially due to the fact that when it comes to infrastructure, the Roman road network, dating back about 2000 years, represents one of the largest infrastructural investments in history.

The literature on the impact of ancient infrastructures refers not only to the Roman roads, but also to the colonial railways or even to the ancient ports, while maintaining one point in common: almost all the works focus on urban growth (Jedwab

and Moradi, 2016; Jedwab et al., 2017; Brata, 2017; Berger and Enflo, 2017) or economic development (Wahl, 2017; Dalgaard et al., 2018), with one exception being De Benedictis et al. (2018).
This analysis differs from this previous literature by investigating the relationship between ancient Roman infrastructure and the propensity to export.
This work is however closely related to an important previous work by De Benedictis et al. (2018), where the relationship between current differences in trade costs and the historical Roman road network is assessed. Nevertheless, this analysis differs from De Benedictis et al. (2018) in two ways. First, this work has as its main object the propensity to export, instead of trade costs and, second, in this analysis the interconnection of the Italian economy is placed under scrutiny rather than international trade.

Moreover, this study contribute to the ongoing debate on the role that public investment should play in the long-term growth process, the focus of which is on the effectiveness of infrastructure investment. In particular, the impact that infrastructures have on production represents the heart of all the literature that has developed around the macro theme of the effectiveness of infrastructure investments. Many works have investigated this issue, as documented by important surveys (Afraz et al., 2006; Ferrari et al., 2018) and meta-analysis (Melo et al., 2013; Bom and Ligthart, 2014).
Production, however, is not the only economic output on which the studies focused; indeed, other outcomes, such as productivity and trade, also found the attention of researchers.
This analysis aims to contribute to this last strand, and in particular to the literature on the impact of roads on trade. Seminal work in this thread of research is Duranton et al. (2014). Indeed, if the previous literature on trade had focused primarily on transport costs and trade barriers, the authors are among the first to investigate the role of infrastructure.

The first part of this analysis is devoted to investigate the long-term impact of the Roman road network on today's propensity to export, while the second part analyzes the possible mechanisms of transmission of this legacy.

To assess the relation between trade and propensity to export, this analysis is based on a micro funded model of roads and trade proposed by Duranton et al. (2014). In particular, it relies on a two-step approach, in which, a standard gravity equation for interregional exports is first estimated in order to recover the fixed regional effects as a proxy for the propensity to export. The regional propensity to export becomes the dependent variable of the second step where the latter is regressed on the endowment of regional Roman roads, controlling for other regional characteristics.
The analysis is based on Ordinary Least Square (OLS) estimates since Roman roads, conditionally on a set of controls, can be considered exogenous, as argued by various

studies (e.g. Percoco, 2016; De Benedictis et al., 2018; Garcia-López et al., 2015).

This analysis refers to the year 2010 and uses data on the value of trade flows between the Italian NUTS-2 regions provided by Thissen et al. (2013).
Different estimates of the famous "gravity with gravitas" model of Anderson and Van Wincoop (2003) are presented in order to retrieve regional fixed effects and then to investigate the long-term impact of Roman road network on export propensity.
The empirical results suggest that Roman roads still have a persistet influence in determining the propensity to export within the national borders, i.e. short-distance trade.

In the second part of this study the mechanisms that could link the ancient road network to the present are investigated. In this part of the work two possible transmission mechanisms are analysed.
First, since the relationship between Roman roads and today's outputs has been explained by a strong relationship with the current road network in various works (e.g. Wahl, 2017; Dalgaard et al., 2018; De Benedictis et al., 2018), and since Roman roads have been used as an instrument for the current road networks in various analysis (e.g. Percoco, 2016; Garcia-López et al., 2015; Garcia-López, 2019; Holl, 2016; Holl and Mariotti, 2018), this relation is put under scrutiny.
The results confirm that the ancient infrastructure is linked to the design and the structure of today's transport network.
Furthermore, following a recent work by Flueckiger et al. (2019), a second transmission mechanism is assessed. More precisely, Flueckiger et al. (2019) demonstrate that territories that in the past were connected by Roman roads now show similarities along different lines, such as preferences, values and attitudes.
More specifically, since the Roman Empire was a fertile ground for trade, it may have left a legacy in terms of institutions and social capital, that might play a role in fostering trade.
Then, the relation between a social capital index and the Roman road network is investigated. However, the results lead to the conclusion that there is no relationship between today's levels of social capital and the ancient Roman roads.

This paper is organized as follows. In Section 2 a short summary of the related literature is presented. The theoretical model by Duranton et al. (2014) is exposed in Section 3 and, in section 4, the identification strategy is presented. Data are described in Section 5. Then, results are exposed in Section 6. Some robustness analysis are proposed in Section 7, and Section 8 contains conclusions.

2 Related Literature

As already mentioned, this work is mainly related to two strands of literature.
First, it relies on the literature on the long-term impact of transport infrastructure

and in particular the very long-term impact of ancient Roman roads, which represents a recent development of the literature on the persistent nature of history. According to Nunn (2009), the origin of the latter can be traced back to the seminal works of Acemoglu et al. (2001, 2002), La Porta et al. (1997, 1998) and Engerman and Sokoloff (1997, 2002). Starting from these six works, Nunn (2009) provides a very complete overview of the literature on the persistence of history[1]. Almost ten years later, in 2017, Michalopulos and Papaioannou (2017) presented another large and complete survey, reviewing all the contributions in this strand of the literature in three different volumes, dividing them through a geographical method (volume 1 global perspective, volume 2 Asia and Africa and volume 3 Americas and Europe).

Within this important strand, a recent literature has focused on the role played by historical infrastructure on present outcomes.

With regard to ancient road infrastructure, the literature focuses mainly on the persistence of the Roman road network (Wahl, 2017; Dalgaard et al., 2018; De Benedictis et al., 2018).

In Wahl (2017) and Dalgaard et al. (2018) nighttime light intensity is used as a good proxy for today economic development and Roman Roads represent a key element in the model used by authors.

In Wahl (2017), a boundary discontinuity design is used to separate Germany into two different areas, following the Limes wall: the treated part, affected by Roman domination, and the untreated area. The dependent variable of the model is the intensity of the night light in 2009 and the main independent variable is a dummy indicating whether or not a pixel was part of the Roman territory in 200 AD. The author notes that areas dominated by the Romans show higher levels of development than others. Once the importance of Roman heritage in determining levels of development in 2009 is observed, the author shows that this long-term effect can be traced back to the persistence of the Roman road network.

In Dalgaard et al. (2018) the territory of the Roman Empire of the second century AD is divided into cells of size 1 degree of latitude and 1 degree of longitude and Roman roads represent the main independent variable, while dependent variables refer to different measures of the economic activity both in the present and in the past. To measure past economic activity authors use the number of Roman settlements in 500 AD, while for today's activity they propose the level of population density and, as in Wahl (2017), the night brightness. The analysis leads to the conclusion that, first, the Roman road network is a good predictor of the road network today in terms of density, and, second, that Roman roads have a lasting effect on economic activity today.

De Benedictis et al. (2018) confirm a strong correlation between the ancient infrastructures, i.e. the Roman road network, and the modern road system, analyzing the

[1]See also Nunn (2014)

Italian provinces. In their work, authors investigate the long-term impact of Roman roads on transport costs. To do this, both a reduced form and an Instrumental Variable (IV) approach are used. De Benedictis et al. (2018) regress the current transport infrastructure on Roman roads and other controls, finding a positive and significant correlation, and then use a Two-Stage Least Squares (2SLS) to assess how today's roads influence transport costs by instrumenting the current transport system with Roman roads. In this way, authors detect a long-term impact of Roman roads via today's infrastructure.

In addition to the Roman network, other past infrastructure investments have also attracted the interest of researchers, especially the ancient railways. The literature on the persistence of ancient railways investments mainly focus on both the short and the long run effect of colonial infrastructure (e.g. Jedwab and Moradi, 2016; Brata, 2017; Jedwab et al., 2017) on urban growth, with the exception of Berger and Enflo (2017).

Jedwab and Moradi (2016) and Jedwab et al. (2017) carry out a spatial analysis using a spatial discontinuity design together with an IV approach to address endogeneity problems. In particular, Jedwab and Moradi (2016) analyse the impact of colonial railways in Ghana and in Africa as a whole. Authors create a spatial database to analyse the spatial equilibrium raised in the period of the construction of the colonial railroad in Ghana and Africa, i.e. 1901-1931, and to investigate whether the spatial equilibrium is still persistent after the disappearance of the colonial railway. The initial model has three different dependent variables: the standard scores of cocoa production, the rural and urban population. In their equations, the railways are cell dummies for being away from a line by different intervals of kilometres. They first estimate the impact of the old railways in the short term, discovering that rail connectivity in 1918 had a strong positive effect on cocoa production in 1927 and on the population in 1931, both rural and urban. Then, authors estimate the effect of rails on the urban population in 2000 and confirm their hypothesis of persistence, demonstrating that, despite the disappearance of the colonial railway system after the independence of Ghana, their effect has persisted in the long-run.

In addition, in the study by Jedwab et al. (2017), the evidence of the persistence of a spatial equilibrium in Kenya is presented. To do so, the authors, after estimating the impact of colonial railways on the past spatial equilibrium, turn to test the hypothesis of persistence using three shocks in the post-independence period. In their main specification, the authors analyse the effect of the 1962 railway connectivity on the population (European, urban, Asian and African) in the same period. They find a positive effect of the presence of railways on all the populations analysed. After independence, in Kenya there was an exodus of settlers and the railway began to decline due to lack of maintenance and because new investment policies aimed at building new sealed roads. Thus, as the authors point out, the equilibrium was

achieved thanks to these factors and, once they disappeared, a very different equilibrium could have emerged after independence. Their study leads to the conclusion that the spatial equilibrium persisted even after independence, since the distribution of the urban population in 2009 is highly correlated to the distribution in the years of independence.
In a very different historical context, that of 19th century in Sweden, Berger and Enflo (2017) analyse the short and long-term impact of a first wave of railway construction, comparing cities that obtained access to the rail network in the period 1855-1870 to cities that, on the contrary, did not. They use a strategy of difference in difference, augmented with an IV approach, instrumenting existing trunk lines with two previous proposals of construction, dating back to 1845 and 1853/54. They observe the impact of the new rail network on the population of each city and find a substantial causal impact of railways on urban growth. After these initial results, the authors investigate another important issue: growth may have been driven by a move from a nearby-unconnected city. Authors find that the increase in population in the cities affected by the rail link was at the expense of other neighbouring cities that did not get access to the network. Analysing the impact of the expansion of the rail network after the first wave of railways investment, what emerges is that this last connection had little impact on the city population and this lead the authors to analyse the potential long-term impact of the first network. To examine the long-term impact, they use ten-year data, from 1800 to 2010, and discover that, during the twentieth century, the population differences between the cities affected by the first wave of railway construction and the "controlling" cities remain significant in every decade from 1860 to 2010.
Following the path of Berger and Enflo (2017), Brata (2017) creates a fictional variable of railway accessibility in the 30s, when the author identifies the end of the construction of colonial railways in Java. The analysis is conducted at the city level, observing the population from 1930 to 2010 and the author concludes that the construction of the railway by the Dutch colonial government had a positive impact in the short term, but does not find evidence in favor of a long-term effect.

In addition to this literature on the importance of the ancient railways, the long-term effect of other modes of transport have been investigated.
In particular, in a recent paper by Jia (2014) the focus is on the *Chinese Treaty ports*[2]. The author investigates the long-term effect of this series of treaties signed by the Quing government, in China, with Western countries, from the 1840s and 1910s. Thanks to these agreements, some Chinese ports were opened to trade by the so-called "unequal treaties", but in 1943 the port system of the treaties ceased to exist. The author proposes both a short and a long term analysis, using a strategy of difference in difference and also an IV approach, using the existence of historical

[2]This expression indicates the port cities of China open to foreign trade during the 19^{th} century.

custom stations as an instrument for *Treaty ports*. The analysis is carried out at prefecture level, dividing between prefectures with *Treaty ports* and control groups without *Treaty ports*. Both a short-term and a long-term impact on the size of the population after the opening of the ports are detected. The author finds that in the short-term, Chinese prefectures that, thanks to *Treaty ports*, had opened up to Western institutions and foreign trade, experienced higher population growth rates. The growth rates between the treated group and the controls aligned after the closure to foreign influence due to the Communist revolution, but in the long run, after 1980, the author observes higher growth rates of GDP and population in the prefectures that were affected by the *Treaty ports*.

This article is related to a second strand of literature, that of the relation between roads infrastructure and trade (Duranton et al., 2014; Duranton, 2015; Volpe Martincus et al., 2017; Coşar and Demir, 2016).[3]

One of the first attempts to assess the impact of roads on trade is Duranton et al. (2014). In their work, authors identify a causal effect of roads on bilateral trade between U.S. cities, using a model inspired by Anderson and Van Wincoop (2003). Their model leads to an estimation strategy very similar to that used by Redding and Venables (2004) with a two-step approach. They first estimate a model of gravity for trade (expressed both in value and in weight) and, on the basis of these results, regress the fixed effects of the exporter and the fixed effects of the importer (respectively the propensity to export and the propensity to import) on the city's highway endowment and other characteristics of the city. To address endogeneity issues, they use three instruments for the current highways network. They rely both on the *"historical route approach"*, using both the routes of major expeditions of explorations, between 1528 and 1850, and the railroad routes in 1898; and on the *"planned route approach"*, using the 1947 plan for the interstate highway system.
Their main finding is a large effect of motorways within cities on the weight of their exports (they estimate an elasticity of about 0.5), but a small effect when they turn to analyse the export value. Nevertheless, a positive and significant impact on export in value is found when the authors analyze short distance trade.

As noted by Duranton et al. (2014), previous trade literature had paid less attention to infrastructure and much more attention to trade barriers and transport costs. The only related work was Michaels (2008). In his research Michaels (2008) estimates the effect of reducing trade barriers on silk demand by using the advent of the US Interstate Highway System as a natural experiment. Using highways that were mostly built between 1959 and 1975, he finds that highways increased trade and thus increased relative demand for skilled manufacturing workers in skill-rich countries and reduced it elsewhere, in line with the Heckscher-Ohlin model predictions.

[3] Also the impact of others modes of transport on trade have been analysed by researchers: ports (Clark et al., 2004; Blonigen and Wilson, 2008; Bottasso et al., 2018) and airports (Alderighi and Gaggero, 2017).

In 2015, Duranton replicates the same methodology used in Duranton et al. (2014) in a different geographic area, Colombia, to assess the external validity of the results for the United States. The author finds a positive and significant impact of within-city roads on trade.
Moreover, two other important works have analysed the causal relationship between roads and trade, namely Coşar and Demir (2016) and Volpe Martincus et al. (2017). The latter uses a difference in difference strategy, together with an IV approach, to study the impact of an asymmetric expansion of the road network that affected Peru between 2003 and 2010. Following the *"historical route approach"*, they use the pre-Columbian Inca road network as an instrument for the modern changes in the road network. While in Coşar and Demir (2016), the focus is not on a low-income country, but on a high-income country, i.e. Turkey. Authors concentrate on the 2000s, when a huge public infrastructure investment affected the country, with the aim of improving the road network by expanding individual carriageways into dual carriageways. Authors use a standard gravity equation where they regress the trade flows on distances between provinces and gateway provinces and on the increase in road capacity, plus a set of gateways and provinces fixed effect. Moreover, in some specifications, they instrument the change in road capacity with the initial share of expressways. The dependent variable used by Volpe Martincus et al. (2017) is the export of companies, while in Coşar and Demir (2016) the analysis focuses on the foreign trade of the provinces. Overall, both studies show a positive impact of the improvement of the road network on trade.

3 Theoretical Model

The empirical analysis has been implemented by following the theoretical model proposed by Duranton et al. (2014), which has been estimated in large US cities. One of the major inference problem of their model is related to possible endogeneity: if roads are constructed to support trade, regions that trade more have a major roads endowment than others or, alternatively, if roads are built to foster trade, region that trade less have higher roads endowment than others. To address this issue authors use an IV approach, using historical and planned networks as instruments for the current highways endowment. Since this study aims to detect if there is a long-term impact of ancient roman roads on the propensity to export of Italian regions, an IV approach is not needed. Indeed there exist historical evidence proving the Roman roads are not affected by endogeneity problems once one control for certain relevant covariates, and several studies confirm this assumption (De Benedictis et al., 2018; Percoco, 2016; Garcia-López, 2012; Garcia-López et al., 2015; Garcia-López, 2019; Holl, 2016; Holl and Mariotti, 2018). This issue will be analysed in depth later, when explaining the identification strategy.

In the model proposed by Duranton et al. (2014), authors developed a theory of roads and trade to derive a characterization of the equilibrium relationship between them.

The starting assumption is that each city produces a unique set of differentiated goods as in Anderson and Van Wincoop (2003). The authors analyse a finite sample of cities they index by $i \in \{1,...,I\}$, and in each city i there are $N_i \in \mathbb{P}_{++}$ identical consumers, each of whom both purchases goods and supplies a unit of labor in city i. Moreover, they consider a continuum of sectors $k \in [0,1]$ to account for the fact that city employment in most sectors is small relative to total city employment. Firm's production function is defined as follows:

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

where A_i represents the productivity of city i, while N_i^k is the employment in city i' s sector k, and it is endogenously determined. Assuming that consumers in all cities have identical preferences, a consumer in city j chooses consumption q_{ij}^k in order to maximise the following utility function:

$$U_j = \Big[\sum_{i=1}^{I} \int_0^1 (q_{ij}^k)^{\frac{\sigma-1}{\sigma}} dk\Big]^{\frac{\sigma-1}{\sigma}}, \tag{2}$$

where σ is the elasticity of substitution between goods and it is assumed to be greater than 1. The consumer's budget constraint is:

$$W_j = \sum_{i=1}^{I} \int_0^1 P_{ij}^k q_{ij}^k dk, \tag{3}$$

in which W_j represents the unit wage in city j and the term P_{ij}^k is the price of the variety from city i in sector k that consumers face in j, and it indicates that price of that variety is different when consumed in city j and not in city i, where it is produced. At this point, the authors introduce the key term of the model, that is X_{ij}^k, the value of pairwise trade in a particular good. To obtain the value of this term it is necessary to maximise utility of Equation 2 subject to the budget constraint in Equation 3, aggregating across consumers:

$$X_{ij} = P_{ij}^k Q_{ij}^k = \Big(\frac{\mathbb{P}_j}{P_{ij}^k}\Big)^{(\sigma-1)} N_i N_j. \tag{4}$$

In Equation 4, $\mathbb{P}_j$ represents the price index of city j and in formula it is defined as

$$\mathbb{P}_j = \Big[\sum_{i=1}^{I} \int_0^1 (P_{ij}^k)^{1-\sigma} dk\Big]^{\frac{1}{1-\sigma}}. \tag{5}$$

At this point, to derive the equilibrium quantity two more steps have to be done: inserting transport costs and deriving the supply relationship. The starting assumption is that the exporter must ship $t_{ij}^k \geq 1$. Thus, the cost to produce the variety from sector k and city i, is multiplied by transport costs at destination j. They obtain:

$$P_{ij}^k = t_{ij}^k \frac{W_i}{A_i}. \tag{6}$$

Finally, the authors arrive to define the equilibrium shipment from city i to city j as follows:

$$X_{ij} = \left(\frac{A_i}{W_i}\right)^{(\sigma-1)} \left[\int_0^1 (t_{ij}^k)^{(1-\sigma)} dk\right] \mathbb{P}_j^{(\sigma-1)} N_j W_j. \tag{7}$$

To derive this equation they insert Equation 6 into Equation 4 and they aggregate pairwise sector flows across sectors. The equilibrium for city i is such that its labor market clears. From Equation 1, it is possible to define the demand for labor as $N_i^k = \frac{Q_i^k}{A_i}$. After some manipulations, they define employment in each city sector:

$$N_i^k = \frac{A_i^{\sigma-1}}{W_i^{\sigma}} \sum_{i=1}^{I} \frac{\mathbb{P}_j^{\sigma-1}}{(t_{ij}^k)^{\sigma-1}} N_j W_j, \tag{8}$$

and aggregate labor supply in city i:

$$N_i^k = \frac{A_i^{\sigma-1}}{W_i^{\sigma}} \Big(\sum_{i=1}^{I} N_j W_j \int_0^1 \frac{\mathbb{P}_j^{\sigma-1}}{t_{ij}^k} dk\Big). \tag{9}$$

As they notice, knowing population in this city, labour productivity and transportation costs in all sectors, Equation 9 yields the wage in city i, so that they can use this wage to compute the value of exports from city i to city j in any sector k.

So far, the model does not consider the role of roads endowment in city i or city j. Roads are supposed to play an important role in the determination of transport costs. Authors, to introduce this relationship in their model define the following measures: R_i, that represents the road capacity of city i, i.e. the kilometres of interstate highways within city boundaries, and R_j, that is the same measure but referred to city j, and R_{ij}, that represents the roads connecting i to j, i.e. the distance measured in kilometres of highways network. Transport cost differ across sector and then, defining V^k as the weight of one unit of output of sector k, they rank sectors from 0, that produces the lightest good, to 1, that produces the heaviest good: $V^0 \leq ... \leq V^k \leq ... \leq V^1$. The authors decompose transport costs in three components:

- the cost of leaving city i, $t_x^k(R_i)$, or more simply t_i^k, that decreases with an increase in its roads endowment, so that $\frac{\partial t_x^k(R_i)}{\partial \log(R_i)} < 0$;
- the cost of entering city j, $t_m(R_j)$, or more simply t_m, that decreases with an increase in its roads endowment, so that $\frac{\partial t_m(R_j)}{\partial \log(R_j)} < 0$;
- the cost of transportation between the two cities, t_{xm} (R_{ij}), or more simply $t_x m$,

that increases with an increase in the distance between the two cities, so that $\frac{\partial t_{xm}(R_{ij})}{\partial \log(R_{ij})} > 0$.

The total transport costs can be expressed in the following way:

$$t_{ij}^k \equiv t^k(R_i, R_{ij}, R_j) = t_x^k(R_i) t_{xm}(R_{ij}) t_m(R_j) \tag{10}$$

Authors can turn to a characterisation of the equilibrium relationship between roads and trade and they formulate some testable predictions[4]. To obtain the equilibrium between roads and trade it is necessary to insert Equation 10 into Equation 7 and, then, to take the logs:

$$\log X_{ij} = \log\Big\{\Big(\frac{A_i}{W_i}\Big)^{(\sigma-1)}\Big[\int_0^1 (t_{ij}^k)^{(1-\sigma)} dk\Big]\mathbb{P}_j^{(\sigma-1)} N_j W_j\Big\} =$$

$$= (\sigma-1)[\log(A_i) - \log(W_i)] + \log\int_0^1 (t_i^k)^(1-\sigma)\log t_{ij} +$$

$$+(\sigma-1)(\mathbb{P}_j - \log t_j) + log(N_j W_j) =$$

$$= \delta_i^X + (1-\sigma)\log t_{ij} + \delta_j^M \tag{11}$$

In Equation 11 two terms appear:

- δ_i^X is the propensity to export, or the exporter fixed effect;
- δ_j^M is the propensity to import, or the importer fixed effect.

From Equation 11 the authors derive the first prediction: *"a reduction in road distance between city i and city j increases the value of trade between these two cities but does not affect its composition"*. The proof of this prediction is straightforward, inserting Equation 6 into Equation 4 it is possible to show that the ratio of export between any two sectors, $\frac{X_{ij}^{k'}}{X_{ij}^k}$, does not depend on t_{ij}, since for Equation 10 transportation costs between city i and city j are the same for all sectors:

$$\frac{X_{ij}^{k'}}{X_{ij}^k} = \frac{P_{ij}^{k'} Q_{ij}^{k'}}{P_{ij}^k Q_{ij}^k} = \frac{t_{ij}^{k'} \frac{W_i}{A_i} Q_{ij}^{k'}}{t_{ij}^k \frac{W_i}{A_i} Q_{ij}^k} = \frac{(t_x^{k'} t_{xm} t_m) \frac{W_i}{A_i} Q_{ij}^{k'}}{(t_x^k t_{xm} t_m) \frac{W_i}{A_i} Q_{ij}^k} = \frac{t_x^{k'} Q_{ij}^{k'}}{t_x^k Q_{ij}^k} \tag{12}$$

Before turning to the second prediction, it is necessary to show few more steps to derive a new measure, that plays a central role in the empirical model, the export market access term. From Equation 11 , the propensity to export can be expressed as:

$$\delta_i^X = \log\Big(\frac{A_i}{W_i}\Big)^{\sigma-1} + \log\int_0^1 (t_i^k)^{1-\sigma} dk, \tag{13}$$

[4] For the purpose of this work, only some of the testable predictions are analyzed

and using Equation 9:

$$\delta_i^X = \log(N_i W_i) - \log \sum_{i=1}^{I} \frac{\mathbb{P}_j^{\sigma-1}}{(t_{xm} t_m)^{\sigma-1}} N_j W_j. \tag{14}$$

The market access term is given by:

$$MA_i^X \equiv log \sum_{i=1}^{I} \frac{\mathbb{P}_j^{\sigma-1}}{(t_{xm} t_m)^{\sigma-1}} N_j W_j = \log \sum_{i=1}^{I} e^{(1-\sigma)\log t_{ij} + \delta_j^M}. \tag{15}$$

Using Equation 9 to eliminate the endogenous wage, W_i, from Equation 14 and defining the function of within-city roads as follows:

$$S(R_i) \equiv \frac{1}{\sigma} \log \int_0^1 (t_i^k)^{1-\sigma} dk, \tag{16}$$

Equation 14 can be rewritten as:

$$\delta_i^X = S(R_i) + \frac{\sigma-1}{\sigma} \log A_i + \frac{\sigma-1}{\sigma} \log N_i - \frac{\sigma-1}{\sigma} MA_i^X. \tag{17}$$

Once defined the market access term, the authors can derive the following testable prediction: *"Export market potential, MA_i^X, negatively affects city i's propensity to export, δ_i^X"*. This directly comes from Equation 17 but this counter-intuitive prediction, as the authors point out, comes from the fact that Equation 11 that describes δ_i^X, already account for distance to importers, $(1-\sigma)\log t_{ij}$, and their propensity to import, δ_j^M. In this way, Equation 17 only captures an indirect wage effect. Indeed, by Equation 7, higher wages reduce the propensity to export, and by Equation 9 wages and market access are positively related, then, the negative effect of market potential on propensity to export is obtained. The third prediction is the following: *"The effect of within-city roads on the propensity to export value should be positive but small"*. Authors, indeed, show that the model allows roads to affect the value of export only through the effect that they have on wages. From Equation 17 is possible to see how this is possible: the term $S(R_i)$ only appear in this equation when Equation 9 is used to eliminate the endogenous wage. From Equation 9, indeed, the log of wage can be described as follows:

$$\log W_i = \frac{\sigma-1}{\sigma} \log A_i + \frac{1}{\sigma} \log N_i + \frac{1}{\sigma} MA_i^X + S(R_i). \tag{18}$$

4 Identification Strategy

4.1 First-step: the gravity equation

The model presented by Duranton et al. (2014) leads to a two-step estimation. From Equation 11 it is derived the first stage of their econometric model:

$$\log X_{ij} = \delta_i + O_l(\log Distance_{ij}) + \delta_j + \epsilon_{ij}. \quad (19)$$

To estimate Equation 19 the authors allow an approximation of transportation cost between city i and city j in the following way:

$$(1-\sigma)\log t_{ij} \equiv O_l(\log Distance_{ij}) + \epsilon_{ij}, \quad (20)$$

where $O_l(logDistance_{ij})$ is an order l polynomial in highways distance between cities. Equation 19 describes a "gravity" equation for trade with fixed effects. The error term, ϵ_{ij}, includes two different components: first, it includes approximation error since $(1-\sigma)\log t_{ij}$ has been substituted with a polynomial distance between i and j; second, it includes transportation costs between i and j that are not related to distance.

The model from Duranton et al. (2014) relates to trade between large US metropolitan statistical areas, but Italian data on bilateral internal export refers only to interregional trade (NUTS2) and there are no data on province level export flows[5]. To deal with this substantial difference between the unit of analysis of this work and that used by the authors, some arrangements are needed. If one could observe bilateral trade between all the 110 Italian provinces existing in 2010, i.e. $logX_{ij}$, it would be possible to estimate Equation 19, but a different quantity is observed: the aggregate export between regions. The latter can be thought as the sum of all bilateral trade between the provinces belonging to an exporting region with the provinces of an importing region. More in general, if exports between provinces are observed at any aggregate level Equation 19 needs to be rewritten as follows:

$$\sum_{i=1}^{n_E}\sum_{j=1}^{n_I} \log X_{ij} = n_I \sum_{i=1}^{n_E} \delta_i + \sum_{i=1}^{n_E}\sum_{j=1}^{n_I} O_l(\log Distance_{ij}) + n_E \sum_{j=1}^{n_I} \delta_j + \sum_{i=1}^{n_E}\sum_{j=1}^{n_I} \epsilon_{ij}, \quad (21)$$

where i represents the exporting province and j the importing province, and n_E is the total number of exporting provinces, while n_I represents the total number of importing provinces.

The best way to conduce the analysis at an aggregate level is to rearrange the model in terms of average values, and after some calculations the resulting equation takes

[5] In the Italian administrative setting regions correspond to NUTS-2 level and provinces correspond to NUTS-3 level.

the following form[6]:

$$\log \overline{X}_{ij} = \frac{\sum_{i=1}^{n_I} \delta_i}{n_E} + \frac{\sum_{i=1}^{n_E} \sum_{j=1}^{n_I} O_l(\log Distance_{ij})}{n_E n_I} + \frac{\sum_{j=1}^{n_E} \delta_j}{n_I} + \frac{\sum_{i=1}^{n_E} \sum_{j=1}^{n_I} v_{ij}}{n_E n_I}, \tag{22}$$

where:

- $\overline{X}_{ij}$ is the average value of export between provinces, i.e. the aggregate average export,
- $\frac{\sum_{i=1}^{n_I} \delta_i}{n_E}$ represents the average fixed effect of the exporting provinces and, reversely,
- $\frac{\sum_{j=1}^{n_E} \delta_j}{n_I}$ represents the average fixed effect of the importing provinces;
- $\frac{\sum_{i=1}^{n_E} \sum_{j=1}^{n_I} O_l(\log Distance_{ij})}{n_E n_I}$ is obtained by calculating distances between each pair of provinces and taking the average value, and
- $\frac{\sum_{i=1}^{n_E} \sum_{j=1}^{n_I} v_{ij}}{n_E n_I}$ is an error term.

In principle, Equation 22 makes it possible to calculate average values at any level of aggregation. In this work the aggregate value of exports observed is the regional value, therefore allowing the exporting provinces i to belong to the same NUTS-2 region and so do the importing provinces j, Equation 22 can be rewritten in terms of regional average values. The resulting equation estimated for the first-step of the model takes the following form:

$$logX_{mn} = \varphi_m + O_l(logDistance_{mn}) + \varphi_n + v_{mn}, \tag{23}$$

where m is the export region, with $m = 1,...,21$, and n is the import region, with $n = 1,...,21$. In Equation 23 all the variables, i.e. export and distances, represent regional averages calculated according to Equation 22.

4.2 Second-step: the impact of Roman roads

Equation 17 suggests to regress the estimated propensity to export $\hat{\varphi}_m$ from the first step on roads endowment, productivity, population and market access.
Estimated equation for propensity to export will be:

$$\hat{\varphi}_m = \beta_0 + \gamma \log RomanRoads_m + \beta C_m + \mu_m \tag{24}$$

where both the Roman roads and the other controls, C_m, are expressed in regional averages, i.e. taking their aggregate regional values divided by the number of provinces

[6]A formal proof of this result is given in Appendix A

of the exporting region, n_E. This adjustment was necessary for reasons of consistency with Equation 23 where φ_m and φ_n should be interpreted as average fixed effects, respectively for exporters and importers.

Since in this model the aim is to give a causal interpretation to the parameter of Roman roads on the propensity to export, an important matter is that of possible endogeneity. This issue has been analysed by several works that have used Roman roads as valid instruments (e.g. Garcia-López et al., 2015; Garcia-López, 2019; Percoco, 2016; Holl, 2016; Holl and Mariotti, 2018; De Benedictis et al., 2018). Overall, the Roman road network have been found exogenous with respect to various economic outcomes. Indeed, since the validity of an instrument hinges on its relevance and on its orthogonality conditional on controls, the several uses of the Roman road system as an instrument confirm the requirement of exogeneity. In particular, as argued by Dalgaard et al. (2018), Roman roads are strongly predetermined and, more in general, almost any ancient transport network may be exogenous because of time that has elapsed since it was built (Duranton and Turner, 2012). The literature has identified *military reason* as the main purpose of Roman road construction, thus excluding a direct economic reason for their location (e.g. Garcia-López et al., 2015; De Benedictis et al., 2018). Moreover, Dalgaard et al. (2018) report two other reasons, besides the military one, why the assignment of the Roman road can be thought of as a natural experiment: the preferred *straightness* of construction and their creation in *newly conquered* and often undeveloped regions.

However, since geography may have influenced the construction of Roman roads, the exogeneity of this ancient network is also based on the use of geographical controls (e.g. Garcia-López, 2019)[7], the percentage of mountainous territory in each region is included as control.

It is worth noting that, De Benedictis et al. (2018) have recently analyzed in depth the role of earlier settlements in determining the location of Roman roads to assess their exogeneity. Specifically, they analyze the correlation between the Roman roads and the presence of cities or settlements that existed before their construction. The authors base the analysis on two different measures. First, they consider the history of Italian provincial capitals to observe which of them was an important urban agglomeration even before the Roman domination. Second, considering all the current provinces, they analyze whether there were pre-Roman amenities throughout the territory. By constructing the relative dummy variables, "Pre-Roman City"

[7] De Benedictis et al. (2018) conduct a specific analysis on this issue estimating an equation where the kilometers of Roman roads, differentiating them between *major* and *minor*, are regressed on an elevation index and on the percentage of mountainous territory. The results suggest that: i) a greater number of kilometres of Roman roads, both *major* and *minor*, are necessary to reach more difficult territories and ii) *major* Roman roads are less dense in more inaccessible territories. The authors conclude that that ancient roads, like the modern ones, are less dense in areas where construction would have entailed higher costs, but the role of geography in determining the location of Roman roads has been limited, i.e. geography is not a key determinant of their construction.

and "Pre-Roman Amenities", that assume value 1 in case pre-Roman agglomerations were found and 0 otherwise, the authors estimate a regression in which the length of Roman roads is regressed on such dummies, controlling also for geographical factors. What emerges in their analysis is that the role of archaic urban agglomeration is minimal for both *major* and *minor* roads. Only in one specification, by analysing the possible role of pre-Roman amenities they find a correlation with Roman roads when they include even *minor* roads in the definition of Roman roads.
Despite this last finding, this analysis relies on a single measure of Roman roads, by aggregating the minors and the majors, for the following reasons. The authors first present estimates by using *major* roads, and to extend the model to *minor* roads, they replicate their model including *all* Roman roads, finding no discrepancies between the two estimates. In addition, including *all* Roman roads in the model is considered necessary for two reasons. First, since the second-step regression is estimated with a sample of 21 observations, the Italian regions, using only *major* roads would have left out of our analysis one region, i.e. Sardinia. Second, since in the second-step relies on average values of Roman roads and not all Provinces are crossed by *major* roads, it seems more accurate to assign regional average values based on both *major* and *minor* roads since the latter widespread throughout the country.

5 Data

5.1 Gravity model database

5.1.1 Distances

The starting database to estimate the gravity equation contains the distances between the centroids of each of the Italian provinces, and then the average *road* distance between the provinces of each region and the provinces of the other regions was computed. The distance was specifically calculated by "`osrm package`" in R that get the shortest path between two points. The first stage database has 441 observations, i.e. the internal trade in each region is also included since it represents how much on average the provinces of a region trade with each other. As alternative measure of distance, it was used the *geodetic* distance between provinces, i.e. the length of the shortest curve between two points along the surface of a mathematical model of the earth.
The Italian regions are considered to be 21, accounting for the fact that the data for two autonomous provinces are reported separately. For this reason, 21 exporter fixed effects are computed, instead of 20, the actual number of regions in Italy.

5.1.2 Export value

The dependent variable of the gravity model is the value of bilateral trade flow between Italian regions in 2010. To obtain data on export in value the database used is from Thissen et al. (2013), a database available online in the Smart Specialisation Platform by the European Commission[8]. This is a unique dataset on bilateral trade flows, expressed in value (millions of euros), between 256 European NUTS-2 regions in 2000, 2005 and 2010. Thissen et al. (2013) use different sources to obtain a final matrix of the most likely trade flows between European regions given all available information.
A potential problem with this data is that trade flows are computed with no differentiation between means of transport and it is not possible to isolate the road transport component. Nevertheless, this analysis use all trade flows in value as a good proxy for trade flows by road because it is documented that the largest part of transport good in Italy is made by road (Ministero delle Infrastrutture e dei Trasporti, 2010-2011)[9].
Table 1 offers some descriptive statistics for the "regional average values" according to Equation (23) of road distance, geodetic distance and value of export.

Table 1: Descriptive Statistics for Main Variables in Gravity Model.

VARIABLES	Obs.	Mean	sd
Value of Export (millions of euros)	441	342.2	2,354
Road Distance (km)	441	605.5	342.8
Geodetic Distance (km)	441	441.5	254.0

Notes: Variables are expressed in terms of "average values" consistent with Equation (23).

It is worth noting that Duranton et al. (2014) and Duranton (2015) use two different trade measures, i.e. trade expressed both in weight and in value, while in this work trade flows between regions are only expressed in value. This choice stems from the fact that although there are data expressed in weight provided by the Italian National Institute of Statistics (ISTAT) in the National Report of Infrastructures and Transports, these data present some measurement characteristics that can lead to biased estimates. Indeed, they refer to all vehicles that have a payload greater than 3.5 tonnes and have been registered in Italy. In this way, two important segments of the transport supply are excluded from the survey. Indeed, the class of vehicles with a payload less than 3.5 tonnes represents a large share of the total number of circulating vehicles and in the last few years, the number of foreign vehicles is increasing

[8]Url: https://s3platform.jrc.ec.europa.eu/s3-trade-tool
[9]More detailed data are presented in Appendix B.

because of the different cost levels of the Italian firms and that of the Eastern Europe (Giordano, 2010).
The main results of Duranton et al. (2014) concern exports by weight, in fact the authors find that the increase in road infrastructure increases the weight of US cities' exports but not their value. Nevertheless, a positive and significant impact on exports in value is found when considering short distance trade. Moreover, in Duranton (2015), where the author analyze the effect of road endowments on trade between Colombian cities, a positive and significant impact is found for trade in value. From these results it emerges therefore that for the short distance trade also the trade in value is influenced by the provision of infrastructure. Since trade between Italian regions is mainly a short distance trade, and considering the possible problems in data on the weight of export, this work concerns trade data expressed only in value.

5.2 Second-step database

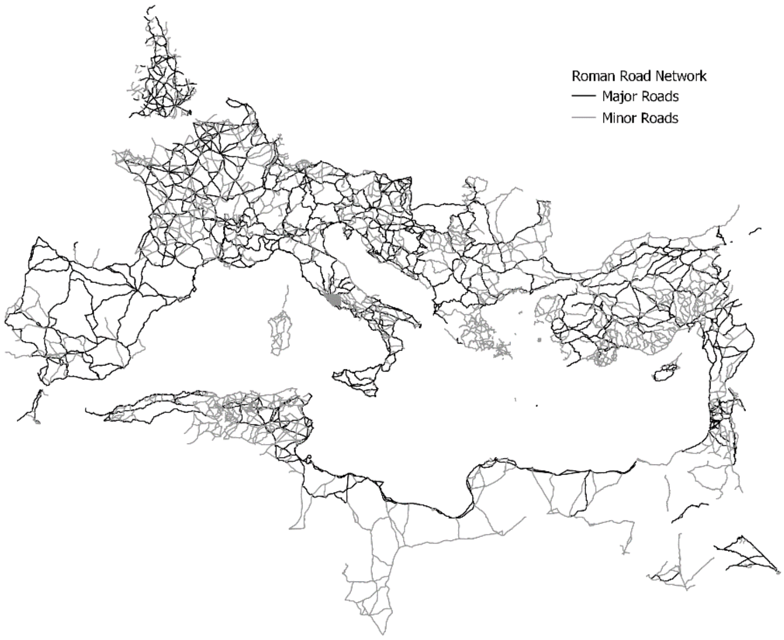
5.2.1 Roman Roads

The dataset containing information on Roman roads at the Italian regional level is constructed on the data provided by McCormick et al. (2013) and available in the Digital Atlas of Roman and Medieval Civilization (DARMC). The DARMC contains geodatabases on the Roman and medieval words and, among other things, it makes information on Roman road network available online. The main predecessor of this database is Talbert (2000) that, with a great cartographic effort, provides maps of the entire world of the Greeks and Romans, covering the territory of more than 75 modern countries. The shapefile contains 7,154 segments representing the Roman roads existing in the Roman Empire in 117 AD. Each segment contains information about its position, its size (*major* or *minor* roads), its certainty and its length expressed in meters. Figure 1 shows the entire Roman road network, differentiating between *major* and *minor* roads, as presented by the DARMC.

By integrating DARMC with the shapefile provided by ISTAT, it was possible to restrict the analysis to the Italian territory and to calculate the length of Roman roads in each Italian province and then to obtain average number of kilometers in each NUTS-2 region. Figure 2 shows the resulting map. In particular, in the Italian peninsula the total length of segments classified as *major* roads is almost 10 kilometres, while, considering *all* roads, i.e. including *minor* roads, the length almost doubles with almost 20 kilometres. The data on Roman roads are presented in the Appendix C[10].

[10]The resulting database is consistent with data presented in a recently published working paper by Licio (2019)

Figure 1: Roman Road Network: Major and Minor Roads.



Source: Author's elaboration from McCormick et al. (2013)

5.2.2 Market Access

Since the estimation strategy is expressed in terms of arithmetic mean (Equation 24), it was necessary to use a different approach to derive the Market Access term. Indeed, using Equation 15 to calculate this term in terms of regional mean appears impossible, since taking the average of that quantity should involve a knowledge of the importer fixed effect at NUTS-3 level, and this information is not available: from Equation 23 only a regional average exporter fixed effect can be estimated. To deal with this issue, an alternative measure proposed by Duranton et al. (2014) in Appendix H is used in this work. The market access term is defined *"as the sum of aggregate income in other cities with the contribution of each city weighted by distance to power -0.9"*, on the basis of Disdier and Head (2008)'s meta analysis where -0.9 represents the mean of all distance decay rates for trade. In this way, using data on GDP from ISTAT, the estimates Market Access for each province is:

$$MA_i = \sum_{j=1}^{n_I} GDP_j (Distance_{ij})^{-0.9} \tag{25}$$

By summing all the Market Access term of each provinces, i, belonging to the same exporting region and dividing for the number of the provinces, n_E, the average regional value was calculated:

$$\overline{MA}_i = MA_m = \frac{\sum_{i=1}^{n_E} MA_i}{n_E}. \tag{26}$$

Figure 2: Roman Road Network in Italy: Major and Minor Roads.

Source: Author's elaboration from McCormick et al. (2013)

5.2.3 Other regional characteristics

In order to collect data on other variables used as controls, the main source is ISTAT. Data on the number of employed persons from 1995 to 2010, as well as, for the same period, those on the manufacturing value added on the total of each region were collected. Since it is necessary to take into account the orographic aspects of the territory, data on the percentage of mountainous territory in the total area of each region are also used. Moreover, to collect information on historical GDP, data refer to information contained in the statistical appendix of Felice (2015), which provides regional information both on GDP per capita and on the share of GDP on the total Italian territory, in the years 1872, 1881, 1891 and 1901. The historical population derives from the summary of historical statistics of Italian National Institute of Statistics (2011), which provides data on the absolute number and percentage of population in 1931, 1936, 1951, 1961, 1971, 1981 and 1991.

Table 2 presents summary statistics for the main variable used in the second-step analysis.

6 Empirical Results

6.1 Gravity model for interregional trade

Table 3 reports OLS estimates for the gravity model (Equation 23). All regressions include exporter fixed effects, which become the dependent variable in the second step of this analysis, and importer fixed effects. The latter capture the inward multilateral resistances, while the exporter fixed effects account for the outward multilateral

Table 2: Descriptive Statistics for Main Variables in Second-step Model

VARIABLES	Obs.	Mean	sd
Roman Roads (km)	21	195.7	160.5
Employment (Thousand People)	21	193.5	102.5
Manufacturing Value added (Million euro	21	1,388	1,033
GDP 2010 (Million euro)	21	13,209	8,482
Surface (km^2)	21	3,201	1,458
%Population 1971	21	0.828	0.427
%GNP on total 1901	21	0.891	0.435
%Mountainous Area	21	21.92	33.44

Notes: Variables are expressed in terms of "average values" consistent with Equation (24).

resistances.

By definition, both exporter and importer fixed effects represent a proxy for export and import propensity and absorb, respectively, the exporter value of output and importer expenditure, as well as all other observable and unobservable exporter - and importer- specific characteristics that may influence bilateral trade (Yotov et al., 2016).

In Table 3 estimates for the elasticity of trade with respect both to travel (or road) distance (Column 1,2, and 3) and to geodetic distance (Column 4,5 and 6) are presented. The elasticity of trade with respect to travel distance is -1.37 (Column 1) and the elasticity is about the same for geodetic distance, i.e. -1.32, (Column 4). Estimates in Column (2) of Table 3 suggest that the effect of distance on trade is negative but such effect is not linear (see also figure 3). These results therefore show a U-shaped relationship between distance and export. Distance would have a positive effect on trade only after about 8297 km and, since average distance between regions is always less than this threshold, this is not the case for the observed sample[11].

Results for export in value are generally coherent with that found by Duranton et al. (2014). In particular they find an elasticity of trade with respect to distance of -1.90 by regressing the value of export on the log of highway distance, and they find an U-shaped relation when introducing the quadratic term (-2.18 for (log) highway distance and 0.062 for the quadratic term).

Estimates in Column (3) refer to a gravity equation where the cubic term is also included. While interpreting the parameters is more involved, using straightforward

[11]The model is also estimated with data in export but as expected this relation is not confirmed. Indeed trade appears to decrease at an increasing marginal rate. This bias is attributed to the fact that data provided by ISTAT ignore the vehicles with a payload less then 3.5 tonnes and foreign operators in calculating trade flows. The data can be considered downward bias: indeed, zero trade flows could not be zero when considering these two important segments of the transport supply. Even when trade flows are not zero they can be underestimated. In particular, it is plausible that this bias affects more trade flows between closer regions.

Table 3: First-step results – Value of Export

	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS
Dependent variable: (log) Value of bilateral trade flows (2010)						
log(Distance)	-1.367***	-2.671***	2.188*			
	(0.0890)	(0.488)	(-1.155)			
log(Distance)^2		0.148***	-1.385***			
		(0.0451)	(0.324)			
log(Distance)^3			0.124***			
			(0.0250)			
log(Geo.Dist)				-1.321***	-3.016***	1.193
				(0.0857)	(0.499)	(-1.014)
log(Geo.Dist)^2					0.196***	-1.190***
					(0.0477)	(0.289)
log(Geo.Dist)^3						0.118***
						(0.0231)
Observations	441	441	441	441	441	441
R-squared	0.767	0.798	0.826	0.754	0.804	0.831

Notes: All regressions include importer and exporter fixed effect for all regions. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

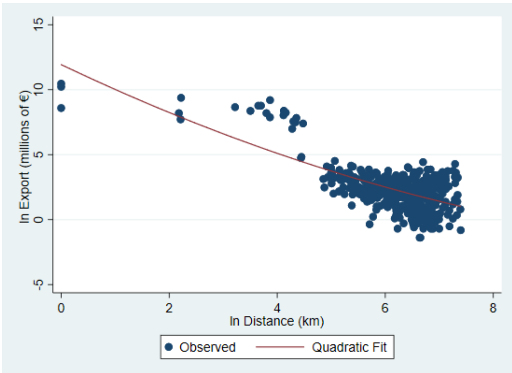


Figure 3: Predicted export from a quadratic regression on distance

calculus it appears that distance continue to have a negative effect on trade[12].

All results are confirmed when the error terms of all coefficient of distance estimated in Table 3 are bootstrapped with a two-way cluster option, i.e. clustering for the exporting and the importing region.

6.2 The effect of Roman Roads on today's propensity to export

All fixed effects estimated through the different specifications of the gravity equations in Table 3 are highly correlated in pairs and, to estimate the equation of the second step, the exporter fixed effects derived from the gravity equation presented in Table 3 Column (2) are used.

The choice is simply linked to economic reasons, because introducing a quadratic

[12]These findings are confirmed plugging in the partial derivative of Equation 19 with respect to distance both the average value and the median value of export.

term can lead to a more accurate approximation of reality, i.e. in this way it is possible to account for a non constant elasticity. Nevertheless, results are robust to using the fixed effects estimated in other columns of Table 3.

Table 4 presents OLS estimates. Coefficient resulting from the the simplest specification, i.e. regressing the propensity to export on the natural logarithm of Roman roads controlling for the regional surface and for the orography, are shown in Column (1). It is worth noting that the percentage of mountainous territory as well as the surface area are necessary for different but both important reasons. Specifically, the first control contributes to the exogeneity of the Roman network since the Roman roads were more sparse in the more impervious territories (De Benedictis et al., 2018) and, although in little part, the geography could have influenced the location of the roads (Garcia-López, 2019). Moreover, the second control is necessary because the measure of Roman roads is expressed in kilometers and including a measure of the size of the territory allows to take into account their density.
In Column (2), the Market Access term, calculated from Equation 26, is added and its coefficient is always negative and significant across all specifications, as predicted by the model of Duranton et al. (2014).
In Columns (3) to (6) estimates of different specification of Equation 24 are presented. In particular, other control variables are included.
First, *Log Employment*, i.e. the logarithm of the average number of employed persons in a region in 2010 (Column 3), and the manufacturing value added (Column 4). As expected, the coefficient of *Log Employment* is positive and significant. Even after controlling for the the manufacturing value added, the coefficient of *Log Roman Roads* is still positive and statistically significant and that of *Market Access* and *Log Roman Roads* are not affected.

Overall, these results suggest that even controlling for geographical features and current economic characteristics (such as Market Access and Employment), the Roman roads have an impact on the propensity to export. In particular, the coefficient is always significant and positive and is stable in the different specifications.

Finally, in Column (5) and (6) controls relative to historical variables are introduced, i.e. the percentage of population in each region in 1971 and the share of GDP of each region in 1901. In particular, 1970 represents the end of the Italian economic boom. During this period, many transformations affected the country, both from an economic and social perspective. In those years the effects of migratory flows have decisively influenced the geographical distribution of the population, in close connection with the evolution of economic events. The flow between the South and the Centre-North reached its maximum intensity in the three-year period 1961-1963, and in 1970 a new maximum was reached, after which a descending phase began in which the intensity of the phenomenon was reduced.
Moreover, the inclusion of GDP at the beginning of the 20th century is intended to

Table 4: Second-step results

	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS
Dependent variable: Exporter Fixed Effects						
Log Roman Roads	0.682***	0.436***	0.347***	0.316*	0.324*	0.353**
	(0.194)	(0.136)	(0.101)	(0.160)	(0.171)	(0.145)
Market Access		-0.0702**	-0.129***	-0.124***	-0.0972***	-0.0981***
		(0.0318)	(0.0193)	(0.0249)	(0.0238)	(0.0237)
Log Employment			0.486***	0.565**	-0.644	-0.410
			(0.0894)	(0.238)	(0.869)	(1.245)
Log Surface	YES	YES	YES	YES	YES	YES
Log %Mountainous Area	YES	YES	YES	YES	YES	YES
Log Manuf. Value added	NO	NO	NO	YES	YES	YES
Log % 1971 population	NO	NO	NO	NO	YES	YES
Log %GNP on total1901	NO	NO	NO	NO	NO	YES
Observations	21	21	21	21	21	21
R-squared	0.415	0.564	0.808	0.809	0.833	0.836

Notes: All regressions include importer and exporter fixed effect for all regions and a constant term.All regressions are estimated with weights given by the number of the provinces in a region. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

provide a snapshot of the economic situation of the Italian regions at the end of the deep changes brought about by the 19th century. Indeed, during the century, various economic and social changes broke the equilibrium of the rural economy. In the northern regions began the slow process of formation of modern industry, starting from the expansion of agriculture, leading to a commercial development. If the process of industrialization was slow in the first half of the century, after the national unity it had a strong development with the formation of the national market. Finally, at the turn of the 19th and 20th centuries, heavy industry had a strong impulse which led to the industrial take-off of the country.

In Column (5) and (6), both the Roman roads and the Market access term are stable and significant, but employment is not significant anymore.

The coefficient of *Log Roman Roads* remains significant even after the introduction of these two past measures of GDP and population.

The results suggest that the impact of the Roman roads is not influenced either by geographical features or by current or past economic characteristics. Moreover, the estimated coefficients are stable: excluding the first rudimentary specification, in columns (2) to (6), it ranges from 0.44 to 0.35.

6.3 Persistence channels

This analysis suggests that the Roman road system influences the propensity to export today. In this section, the possible mechanisms underlying this result are discussed[13].

In a very recent paper, Flueckiger et al. (2019) explore four different possible channels through which the Roman network could still give shape to the geography of interregional business links[14]. First, the authors prove the persistence in transport network connectivity, i.e. the transport costs in Roman times are reflected in a greater connection of the road network today. This happens even if the transport network of the past is radically different from the present one in terms of layout and technologies, i.e. a greater connection in Roman times is reflected in a better connection today even when new technologies have been made available. Moreover, they show that as the distance measured in terms of the Roman network between any two grid cells increases, so do the differences in terms of production structure, preferences and values. In general, what emerges from their study is that more connected territories in the past, show better connections now, in terms of roads and passenger transport, more similar production facilities, but also more similarity in terms of preferences (e.g. trust, altruism, risk preference and time preferences) or values and attitudes towards work, religion, and politics. The infrastructural connection of the past is not only transformed into an infrastructural connection today, but also into a socio-economic convergence.

In another important study by Wahl (2017), two different channels are considered to favor the persistence of the Roman road network and the effect of Roman legacy on contemporary economic prosperity of the contemporary Germany. First, the author confirms the hypothesis that the persistence of the Roman road network has allowed a denser road network. Second, the author shows that the amenities of the roman roads network have led to higher city growth in the formerly Roman area of Germany.

Moreover, Dalgaard et al. (2018) find that the density of Roman roads is generally a strong predictor of modern road density, although the construction of Roman roads did not follow the rules of the infrastructure planning in contemporary times[15]. This results is confirmed by De Benedictis et al. (2018), who find Roman road network to have a persistent effect on today roads and railway system in Italy.

Furthermore, the link between the two road systems is confirmed by the various works that have used the Roman roads as an instrument for the modern road infrastructure. In particular, in several papers (e.g. Garcia-López et al., 2015; Garcia-López, 2019; Holl, 2016; Holl and Mariotti, 2018) a positive correlation between Roman roads and current Spanish highways has been shown. In Italy, in addition to the

[13]Descriptive statistics for the main variables used to conduct analysis in this section are presented in Appendix D.

[14]To this end, authors divide the territory of modern-day countries that once lay within the borders of the Roman Empire into equally sized grid cells of approximately 56 kilometres. Their analysis include Austria, Belgium, England, France, Germany, Italy, Luxembourg, Netherlands, Portugal, Spain, and Switzerland.

[15]Authors conduct their analysis across the entire area under dominion of the Roman Empire in the year 117 AD.

work by De Benedictis et al. (2018) already mentioned above, an important analysis is by Percoco (2016). In particular, Percoco (2016) investigates whether the position of a motorway exit in a given Italian city attracts companies with a high transport intensity. To address the endogeneity of the localisation of motorway exits, he uses as an instrument the network of Roman roads.

The small number of observations in the second-step analysis does not allow to adopt an IV approach, however, some evidences on the correlation between the density of the Roman network and the current one using data at the NUTS-3 regional level are provided.

In particular, in the spirit of Dalgaard et al. (2018) the link between road density in ancient times and the current one is analyzed. The dependent variable is the length (km) of the motorways in each province, excluding those provinces that do not have motorways. The main explanatory variable is represented by the length (km) of Roman roads, both *major* and *minor*, and then the analysis is conducted only with *major* roads. Some geographic controls are included: latitude, longitude, ruggedness[16] and elevation. Moreover, two other controls are used: a dummy that indicates whether a province has a costal area and a variable that indicates the distance from Rome, the centre of the Roman Empire. Fixed regional effects are included to take into account the institutional differences that may exist between provinces belonging to different regions. The resulting equation is the following:

$$\log Highway_i = \beta_0 + \delta_R + \beta_1 \log RomanRoads_i + X'\beta_2 + \epsilon_i, \tag{27}$$

where $Highway_i$ is a measure for kilometers of highways in 2010 in each province i, δ_R represents the regional fixed effects, $\log RomanRoads_i$ the Roman roads endowment in the past, and X' is a vector of control variables.

Results are presented in Table 5. Estimates of the more parsimonious specification are reported in Column (1) and suggest that an increase of one percent in the Roman road length is associated to an increase in modern motorway length of around 0.26%. This percentage drops to 0.18 when only the *major* Roman roads are considered.

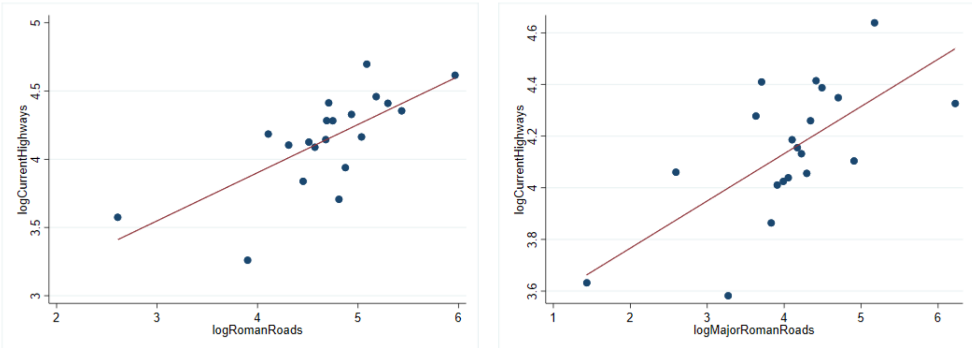
The Roman roads explain about 40% of the variability of the current road network, both considering all Roman roads (Panel A) and limiting the analysis to the *major* roads (Panel B).

When regional fixed effects are included, coefficient for both *all* roads and *major* roads is stable and significant (Column 2).

Moreover, estimates in Column (3) show that none of the geographical variables is significant, while the coefficient of Roman roads, in both panel A and panel B, remains significant. As expected, the coefficient of the distance from Rome is nega-

[16]Source: author's elaboration from Nunn and Puga (2012).

tive and significant (Column 4). When including the dummy on presence or not of a coastal area in each province (Column 5), this coefficient is not significant. Finally, in Figure 4 a binned partial residual plot that refers to the equation estimated in Column (5) of Table 5 is presented. Figure 4 shows the partial correlation between Roman roads and modern roads obtained by estimating Equation 27 and in particular refers to the model estimated in Table 5 Column (5). The positive link between the two road networks appears well determined both for all Roman roads and only *major* Roman roads. Overall, these results confirm the persistent influence of ancient Roman infrastructure on modern infrastructure.



(a) All Roman Roads (b) Only Major Roman Roads

Figure 4: Conditional binned residual scatter plot of the relationship between (log) Highways in 2010 and (log) Roman road for 90 Italian provinces. The underlying equation refers to Table 5 Column (5).

As already mentioned, Flueckiger et al. (2019) investigate different channels through which the Roman transport network can still influence economic integration today and, among other things, they find that history of exchange within the Roman transport network resulted in a convergence in preferences and values. In this spirit, since from an economic point of view in the Roman Empire the role of market forces was central, it is worth investigating whether this centrality can be passed through attitudes. This means imagining that not only do Roman roads have an effect on the propensity to export because they have a strong relationship with today's highway network, ensuring greater ease of exchange, but also that the persistence of Roman road network and the effect of Roman legacy is related to trade attitude associated with social capital and formal and informal institutions. To asses this issue, it is necessary to focus on the factors that have allowed market activities to flourish during the Roman Empire.

According to Temin (2006), the great economic activity developed by the Romans was mainly based both on political institutions and on a legal framework, but they were also favored by social and informal institutions. In particular, informal institutions were a useful tool in addressing problems of incomplete information. Families, large households of slaves and freedmen, and friends were used to reduce the problems of adverse selection and moral hazard: the principals and the agents were part of the

Table 5: Roman Roads and Modern Roads

	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS
Dependent variable: log Current Highways (km)					
PANEL A					
Log Roman Roads	0.262***	0.393***	0.387***	0.365***	0.353***
	(0.0457)	(0.0793)	(0.0827)	(0.0713)	(0.0808)
Log Surface	0.329***	0.388***	0.422***	0.414***	0.434***
	(0.0805)	(0.0871)	(0.0830)	(0.0808)	(0.0949)
Log Ruggedness Index			0.347	0.310	0.375
			(0.462)	(0.449)	(0.524)
Log Longitude			-0.890	-1.032	-1.004
			(1.023)	(0.972)	(1.006)
Log Latitude			5.360	5.699	5.357
			(10.87)	(10.91)	(11.11)
Log Elevation			-0.418	-0.373	-0.434
			(0.517)	(0.503)	(0.575)
Log Distance to Rome				-0.195***	-0.187***
				(0.0502)	(0.0615)
Log Costal Area					0.115
					(0.230)
Region FE	NO	YES	YES	YES	YES
Observations	90	90	90	90	90
R-squared	0.388	0.595	0.609	0.622	0.625
PANEL B					
Log Major Roman Roads	0.183***	0.214***	0.208***	0.196**	0.183**
	(0.0400)	(0.0704)	(0.0716)	(0.0713)	(0.0771)
Log Surface	0.369***	0.484***	0.541***	0.516** *	0.544***
Log Ruggedness			0.481	0.416	0.5047
			(0.525)	(0.489)	(0.570)
Log Longitude			-1.164	-1.338	-1.289
			(0.928)	(0.892)	(0.952)
Log Latitude			4.892	5.583	4.953
			(11.43)	(11.37)	(11.46)
Log Elevation			(11.34)	(11.36)	(11.64)
			(0.591)	(0.552)	(0.628)
Log Distance to Rome				-0.274***	-0.260***
				(0.0542)	(0.0635)
Log Costal Area					0.162
					(0.245)
Region FE	NO	YES	YES	YES	YES
Observations	90	90	90	90	90
R-squared	0.378	0.504	0.528	0.555	0.560

Notes: All regressions estimate the correlation between (log) highaways in 2010 and ancient Roman network for 90 Italian Provinces. All provinces with no higways in 2010 are dropped and 1km of Roman roads, both *minor* and *major*, was added to all provinces to not exclude regions with no *major* roman roads and taking the log. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

same social group and their informal relationships helped the success of business transactions.
The role of formal and informal institutions emerges to be crucial for the Roman mercantile development, and consequently the social capital could represent a possible channel of persistence of Roman road network.

According to Di Liberto and Sideri (2015), first, social capital could influence institutional effectiveness through its effects on the behaviour of bureaucratic elites. Second, if one considers that administrations face agency problems, social capital can positively influence agents' relations with colleagues and supervisors.
Following Di Liberto and Sideri (2015), this possible relationship is assessed by regressing the current social capital on the total length of Roman roads, controlling for past domination:

$$SocialCapital_i = \beta_0 + \beta_1 PastDomininination_i + \beta_2 RomanRoads + X'\gamma + \epsilon_i, \quad (28)$$

where $PastDomininination_i$ is a variable that identifies the number of years in which a given Italian provinces, i, have been dominated by the Normans, the Swabis, the Anjou, the Spanish, the Bourbons, the Savoy, the Austrians and the Republic of Venice[17]. The dependent variable refers to a social capital index built by Cartocci (2007) at the Italian provincial level referring to the early 2000s: it includes indicators of electoral participation, circulation of non-sporting newspapers, blood donations and the importance of sporting activity. Other controls are included in the model: latitude, ruggedness index[18], institutional quality index (average value between 2005 and 2010)[19], past urbanization[20], percentage of religious marriages (2005)[21] and number of extortions (2010)[22].

The results are shown in Table 6. The coefficient of Roman roads is never significantly different from zero, while the coefficient of other past dominations are in line with the results of Di Liberto and Sideri (2015). In particular, estimates confirm the negative effect of Spanish domination and Papal State, as well as the negative effect of Norman and Swabians domination and the positive impact of the Austrians. The only exception is represented by Venice domination whose coefficient is negative and significant. Results in Column(2) confirm the positive relationship between social capital and institutional quality. Furthermore, two geographical indicators, latitude and ruggedness index, are both significant and, respectively, positive and negative (Column 3). Finally, estimates in Column (4) show a positive coefficient for urbanization in 1300, while suggest a negative effect of the number of extortions and the

[17]Source: Di Liberto and Sideri (2015)
[18]Source: authors' elaboration from Nunn and Puga (2012)
[19]Source: Nifo and Vecchione (2014)
[20]Source: Malanima (2016)
[21]Source: ISTAT
[22]Source: ISTAT

Table 6: Roman Roads and Social Capital

	(1) OLS	(2) OLS	(3) OLS	(4) OLS
Dependent variable: Social Capital Index				
Log Roman Roads	-0.0573	-0.0581	-0.0510	0.0614
	(0.289)	(0.280)	(0.287)	(0.210)
Log Surface	0.0293	-0.0310	0.264	0.174
	(0.364)	(0.345)	(0.370)	(0.268)
Normans	-0.0401***	-0.0305***	-0.0200	-0.00383
	(0.0103)	(0.0106)	(0.0144)	(0.0131)
Swabians	-0.00629***	-0.00671***	-0.00924**	-0.00728**
	(0.00199)	(0.00207)	(0.00371)	(0.00365)
Ajou	0613	-0417	-0.00883	0.00519
	(0.00625)	(0.00576)	(0.00690)	(0.00636)
Spain	-0.00851***	-0.00731***	-0.00639**	-0.00333*
	(0.00215)	(0.00219)	(0.00294)	(0.00184)
Bourbons	0.0135	0.0115	0.0184	-0.0170
	(0.0105)	(0.0106)	(0.0135)	(0.0130)
Papal state	-0.00376***	-0.00335***	-0.00277**	-0.00155*
	(0.00111)	(0.00108)	(0.00120)	(0931)
Austria	0.00312*	0.00259*	0.00108	-0836
	(0.00159)	(0.00143)	(0.00202)	(0.00183)
Savoy	-0.00222	-0.00200	-0.00202	-0.00207
	(0.00144)	(0.00140)	(0.00149)	(0.00137)
Venice	-0.00290**	-0.00265**	-0.00500***	-0.00234*
	(0.00131)	(0.00127)	(0.00133)	(0.00138)
Log Latitude			0.362*	
			(0.201)	
Log Ruggedness Index			-0.328***	
			(0.106)	
Log Institution Quality Index		1.068***		
		(0.401)		
Log Urbanisation 1300				0.366***
				(0.128)
% Religious weddings				-0.146***
				(0.0211)
Log Extorsions				-0.674***
				(0.208)
Observations	101	101	101	101
R-squared	0.725	0.759	0.758	0.829

Notes: All regressions estimate the correlation between today's Social Capital Index and ancient Roman network for 101 Italian Provinces out of 103 Provinces. All provinces with no roman roads are dropped. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

percentage of religious weddings.

From these results the social capital does not seem to be a channel of persistence of Roman roads network.

Nevertheless, these results must be treated with caution and this issue deserve a more deeper analysis to obtain conclusive findings.

7 Robustness Analysis

In order to verify the robustness of discussed results, some further analysis has been conducted.

Table 7 presents results of second-step estimates obtained by using different exporter fixed effects.

Estimates for the gravity equation are replicated by not using a quadratic term (Panel A) or by introducing a cubic term (Panel B). The coefficient of (log) Roman roads is still significant and stable in all specifications.

Another issue that deserves some attentions is related to the fact that the dependent variable of the second step is an estimated value. Indeed, since exporter fixed effects are estimated in the first stage regression, it could be necessary to account for the fact that it is not possible to observe their real value, but the real value plus an error term. The best approach to deal with this issue is proposed by Lewis and Linzer (2005).

Table 7: Second-step results, robustness to choice of the gravity equation

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	OLS	OLS	OLS	OLS
Dependent variable: Exporter Fixed Effects from alternative gravity model						
Panel A						
Log Roman Roads	0.715**	0.516***	0.317***	0.301*	0.309*	0.354**
	(0.266)	(0.121)	(0.104)	(0.160)	(0.174)	(0.142)
Observations	21	21	21	21	21	21
R-squared	0.366	0.754	0.857	0.858	0.871	0.876
Panel B						
Log Roman Roads	0.659***	0.581***	0.432***	0.357*	0.365*	0.387**
	(0.152)	(0.114)	(0.107)	(0.167)	(0.180)	(0.158)
Observations	21	21	21	21	21	21
R-squared	0.497	0.611	0.721	0.725	0.759	0.760

Notes: Panel A presents coefficient of log Roman Roads estimated by regressing the exporter fixed effects obtained by equation $\log\overline{X}_{ij} = \varphi_i^{EXP} + \frac{\sum_{ij} O_1(\log R_{ij})}{n_{EXP} n_{IMP}} + \varphi_j^{IMP} + v_{ij},$. Panel B presents the same estimates but exporter fixed effects are obtained from $\log\overline{X}_{ij} = \varphi_i^{EXP} + \frac{\sum_{ij} O_3(\log R_{ij})}{n_{EXP} n_{IMP}} + \varphi_j^{IMP} + v_{ij}$.From Column (1) to Column (6) the controls used are the same as those used for estimating the main model in the Table 4. Robust standard errors in parentheses *** $p < 0.01, **p < 0.05, *p < 0.1$

According to the authors, even using the White et al. (1980) heteroskedastic consistent standard error estimator, the OLS estimator could be inefficient since it does not use an important information about the nature of heteroscedasticity, i.e. the variance of the sampling error. On the other hand, the Weighted Least Squares (WLS) does the opposite: it sets the variance of the error equal to zero and only accounts for the variance of the sampling error in attributing weights. In this way, the WLS estimator can be efficient just if the variance of the homoscedastic noise is very small

Table 8: Second-step results, robustness to choice of the estimation method

	(1) FGLS	(2) FGLS	(3) FGLS	(4) FGLS	(5) FGLS	(6) FGLS
Dependent variable: Exporter Fixed Effects						
Log Roman Roads	0.515***	0.428***	0.289***	0.197*	0.151	0.215*
	(0.145)	(0.138)	(0.0785)	(0.109)	(0.118)	(0.101)
Market Access		-0.0691**	-0.122***	-0.101***	-0.0804***	-0.0845***
		(0.0318)	(0.0196)	(0.0253)	(0.0256)	(0.0210)
Log Employment			0.545***	0.803***	0.0418	0.280
			(0.0888)	(0.236)	(0.643)	(0.634)
Log Surface	YES	YES	YES	YES	YES	YES
Log %Mountainous Area	YES	YES	YES	YES	YES	YES
Log Manuf. Value added	NO	NO	NO	YES	YES	YES
Log % 1971 population	NO	NO	NO	NO	YES	YES
Log %GNP on total1901	NO	NO	NO	NO	NO	YES
Observations	21	21	21	21	21	21
R-squared	0.444	0.564	0.802	0.815	0.836	0.849

Notes: The Table presents coefficient of log Roman Roads estimated by regressing the exporter fixed effects obtained by equation $\log\overline{X}_{ij} = \varphi_i^{EXP} + \frac{\sum_{ij} O_2(\log R_{ij})}{n_{EXP} n_{IMP}} + \varphi_j^{IMP} + v_{ij}$.Robust standard errors in parentheses *** $p < 0.01, **p < 0.05, *p < 0.1$

in comparison to the variance of the sampling error. The larger the share of the variance of the sampling error, the less efficient OLS is and the more efficient WLS is. To overcome this issue, Lewis and Linzer (2005) suggest to use the Feasible Generalised Least Squares (FGLS) estimator that account both for the variance of the sampling error from the first stage regression and for an estimate of the variance of the error in the second stage. Table 8 presents the results for FGLS estimates. The coefficient of Roman roads remains overall stable and significant. Also Market Access and employment do not show substantial variations compared to the results obtained through the OLS estimates.

Finally, overall results remain robust when regions included in the sample change, by removing one observation at a time. The coefficient of Roman roads is stable and significant.

8 Conclusions

This article analyses the long-term impact of Roman infrastructure on the propensity to export of Italian regions.
Conditioning on a set of controls that could explain both the location of Roman roads of 117 AD and today's propensity to trade, it is argued that Roman roads endowment can be considered exogenous. The main results in this paper suggest that Roman roads are positively correlated with regional propensity to trade which has been estimated as suggested by Duranton et al. (2014). In particular, controlling for current economic variables, i.e. employment, market access and value added of manufactur-

ing, and for past variables, i.e. past population and GDP, that could influence the propensity to export, the coefficient of the length of Roman roads is always positive and significant across different specifications suggesting that Roman roads still have a persistent influence in determining the propensity to trade.
A large battery of robustness checks has been performed and results are confirmed when using different fixed effects or using FGLS estimator.

Once the long-term impact of the Roman network is recognized, the question becomes what are the possible channels of persistence.
The first hypothesis, supported in the literature, is that the Roman roads are correlated with today's roads even though they were not established on the same route. The results in this paper support this hypothesis.
Moreover, starting from the assumption that commercial activity was a key point of the Roman Empire, it is investigated a second channel as responsible of the persistence of Roman road network: ancient infrastructure could have left a legacy in terms of trade attitude and institutions, both formal and informal.
To assess this issue a model that correlates a measure of today's social capital with Roman roads, controlling for past dominations between 1100 and 1800 in the Italian peninsula, has been estimated.
However, results in this paper suggest that the Roman roads do not show a relationship with the current share of capital endowment.

References

Acemoglu, D., Johnson, S. and Robinson, J. A. (2001), 'The colonial origins of comparative development: An empirical investigation', *American economic review* **91**(5), 1369–1401.

Acemoglu, D., Johnson, S. and Robinson, J. A. (2002), 'Reversal of fortune: Geography and institutions in the making of the modern world income distribution', *The Quarterly journal of economics* **117**(4), 1231–1294.

Afraz, N., Aquilina, M., Conti, M. and Lilico, A. (2006), 'Impact of transport infrastructure on economic growth', *Annex 6 to Final Report of COMPETE: Analysis of the Contribution of Transport Policies to the Competitiveness of the EU Economy and Comparison with the United States, Karlsruhe, ISI* .

Alderighi, M. and Gaggero, A. A. (2017), 'Fly and trade: Evidence from the italian manufacturing industry', *Economics of transportation* **9**, 51–60.

Anderson, J. E. and Van Wincoop, E. (2003), 'Gravity with gravitas: a solution to the border puzzle', *American economic review* **93**(1), 170–192.

Berger, T. and Enflo, K. (2017), 'Locomotives of local growth: The short-and long-term impact of railroads in sweden', *Journal of Urban Economics* **98**, 124–138.

Blonigen, B. A. and Wilson, W. W. (2008), 'Port efficiency and trade flows', *Review of International Economics* **16**(1), 21–36.

Bom, P. R. and Ligthart, J. E. (2014), 'What have we learned from three decades of research on the productivity of public capital?', *Journal of economic surveys* **28**(5), 889–916.

Bottasso, A., Conti, M., de Sa Porto, P. C., Ferrari, C. and Tei, A. (2018), 'Port infrastructures and trade: empirical evidence from brazil', *Transportation Research Part A: Policy and Practice* **107**, 126–139.

Brata, A. G. (2017), Exploring the Influence of Colonial Railways on Java's Economic Geography, MPRA Paper 80097, University Library of Munich, Germany.
URL: *https://ideas.repec.org/p/pra/mprapa/80097.html*

Cartocci, R. (2007), *Mappe del tesoro: atlante del capitale sociale in Italia*, Vol. 168, Il mulino.

Clark, X., Dollar, D. and Micco, A. (2004), 'Port efficiency, maritime transport costs, and bilateral trade', *Journal of development economics* **75**(2), 417–450.

Coşar, A. K. and Demir, B. (2016), 'Domestic road infrastructure and international trade: Evidence from turkey', *Journal of Development Economics* **118**, 232–244.

Dalgaard, C.-J., Kaarsen, N., Olsson, O. and Selaya, P. (2018), Roman Roads to Prosperity: Persistence and Non-Persistence of Public Goods Provision, CEPR Discussion Papers 12745, C.E.P.R.
URL: *https://ideas.repec.org/p/cpr/ceprdp/12745.html*

De Benedictis, L., Licio, V., Pinna, A. M. et al. (2018), The long-term effects of the historical roman road network: trade costs of italian provinces, Working Paper CRENoS 201801, Centre for North South Economic Research, University of Cagliari and Sassari.

Di Liberto, A. and Sideri, M. (2015), 'Past dominations, current institutions and the italian regional economic performance', *European Journal of Political Economy* **38**, 12–41.

Disdier, A.-C. and Head, K. (2008), 'The puzzling persistence of the distance effect on bilateral trade', *The Review of Economics and statistics* **90**(1), 37–48.

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

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

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

Engerman, S. L. and Sokoloff, K. L. (1997), Factor endowments, institutions, and differential paths of growth among new world economies, *in* S. Haber, ed., 'How Latin America Fell Behind', Standford University Press, Standford, pp. 260–304.

Engerman, S. L. and Sokoloff, K. L. (2002), Factor endowments, inequality, and paths of development among new world economics, NBER Working Paper 9259, National Bureau of Economic Research.
URL: *https://www.nber.org/papers/w9259*

Felice, E. (2015), *Ascesa e declino: storia economica d'Italia*, Il mulino Bologna.

Ferrari, C., Bottasso, A., Conti, M. and Tei, A. (2018), *Economic Role of Transport Infrastructure: Theory and Models*, Elsevier, Amsterdam.

Flueckiger, M., Hornung, E., Larch, M., Ludwig, M. and Mees, A. (2019), Roman transport network connectivity and economic integration, CEPR Discussion Paper DP13838, C.E.P.R.
URL: *https://ideas.repec.org/p/ces/ceswps/*$_7$740*.html*

Garcia-López, M.-A. (2012), 'Urban spatial structure, suburbanization and transportation in barcelona', *Journal of Urban Economics* **72**(2-3), 176–190.

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. and 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.

Giordano, R. (2010), Quello che l'ISTAT non riesce a dire, Trimestrale on line di economia della logistica e dei trasporti. Sistemi di logistica 2.
URL: *http://www.roccogiordanoeditore.eu/files/150326/sdl2010*$_0$2*.pdf*

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

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

Italian National Institute of Statistics (2011), Sommario delle statistiche storiche 1926-1985, Report.

Jedwab, R., Kerby, E. and Moradi, A. (2017), 'History, path dependence and development: Evidence from colonial railways, settlers and cities in kenya', *The Economic Journal* **127**(603), 1467–1494.

Jedwab, R. and Moradi, A. (2016), 'The permanent effects of transportation revolutions in poor countries: evidence from africa', *Review of economics and statistics* **98**(2), 268–284.

Jia, R. (2014), 'The legacies of forced freedom: China's treaty ports', *Review of Economics and Statistics* **96**(4), 596–608.

La Porta, R. L., Lopez-de Silanes, F., Shleifer, A. and Vishny, R. W. (1998), 'Law and finance', *Journal of political economy* **106**(6), 1113–1155.

La Porta, R., Lopez-de Silanes, F., Shleifer, A. and Vishny, R. W. (1997), 'Legal determinants of external finance', *The journal of finance* **52**(3), 1131–1150.

Lewis, J. B. and Linzer, D. A. (2005), 'Estimating regression models in which the dependent variable is based on estimates', *Political analysis* **13**(4), 345–364.

Licio, V. (2019), When history leaves a mark: a new measure of roman roads, Working paper crenos, Centre for North South Economic Research, University of Cagliari and Sassari.

Malanima, P. (2016), 'Italian urban population 1300-1861 (the database)'.

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

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

Michaels, G. (2008), 'The effect of trade on the demand for skill: Evidence from the interstate highway system', *The Review of Economics and Statistics* **90**(4), 683–701.

Michalopulos, S. and Papaioannou, E. (2017), *The Long Economic and Political Shadow of History: Europe and the Americas*, CEPR Press, London.

Ministero delle Infrastrutture e dei Trasporti (2010-2011), *Conto Nazionale delle Infrastrutture e dei Trasporti*, Istituto Poligrafico e Zecca dello Stato S.p.A, Rome.

Nifo, A. and Vecchione, G. (2014), 'Do institutions play a role in skilled migration? the case of italy', *Regional Studies* **48**(10), 1628–1649.

Nunn, N. (2009), 'The importance of history for economic development', *Annu. Rev. Econ.* **1**(1), 65–92.

Nunn, N. (2014), Historical development, *in* 'Handbook of economic growth', Vol. 2, Elsevier, Amsterdam, pp. 347–402.

Nunn, N. and Puga, D. (2012), 'Ruggedness: The blessing of bad geography in africa', *Review of Economics and Statistics* **94**(1), 20–36.

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

Redding, S. and Venables, A. J. (2004), 'Economic geography and international inequality', *Journal of international Economics* **62**(1), 53–82.

Talbert, R. J. (2000), *Barrington Atlas of the Greek and Roman World: Map-by-map Directory*, Vol. 1, Princeton University Press, Princeton.

Temin, P. (2006), 'The economy of the early roman empire', *Journal of Economic Perspectives* **20**(1), 133–151.

Thissen, M., Van Oort, F., Diodato, D. and Ruijs, A. (2013), *Regional competitiveness and smart specialization in Europe: Place-based development in international economic networks*, Edward Elgar Publishing, Cheltnham, UK - Northampton, MA, USA.

Volpe Martincus, C., Carballo, J. and Cusolito, A. (2017), 'Roads, exports and employment: Evidence from a developing country', *Journal of Development Economics* **125**, 21–39.

Wahl, F. (2017), 'Does European development have roman roots? evidence from the german limes', *Journal of Economic Growth* **22**(3), 313–349.

White, H. et al. (1980), 'A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity', *econometrica* **48**(4), 817–838.

Yotov, Y. V., Piermartini, R., Monteiro, J.-A. and Larch, M. (2016), *An advanced guide to trade policy analysis: The structural gravity model*, World Trade Organization Geneva.

Appendix A. Derivation of the first-stage equation

Starting from Equation 21 in the main text, it is possible to rearrange the model in terms of average values as follows:

$$\frac{\sum_{i=1}^{n_E}\sum_{j=1}^{n_I}\log X_{ij}}{n_E n_I} = \frac{\sum_{i=1}^{n_E}\delta_i}{n_E} + \frac{\sum_{i=1}^{n_E}\sum_{j=1}^{n_I} O_l(\log Distance_{ij})}{n_E n_I} + \frac{\sum_{j=1}^{n_I}\delta_j}{n_I} + \frac{\sum_{i=1}^{n_E}\sum_{j=1}^{n_I}\epsilon_{ij}}{n_E n_I}. \tag{A.1}$$

The left-hand side of Equation A.1 is the geometric mean of bilateral trade flows between provinces:

$$\frac{\sum_{i=1}^{n_E}\sum_{j=1}^{n_I}\log X_{ij}}{n_E n_I} = \log\left(\prod_{ij} X_{ij}\right)^{\frac{1}{n_E n_I}} = GM\{X_{ij}\}, \tag{A.2}$$

where the geometric mean is indicated by $GM\{X_{ij}\}$.
Since there is no knowledge about trade flows between provinces, X_{ij}, Equation A.2 cannot be computed. The best approximation of this quantity is the logarithm of the arithmetic mean plus an error term, due to the fact that the geometric mean is substituted by the arithmetic mean. To this end, it is necessary to define: $\gamma_{ij} = \frac{X_{ij}}{AM\{X_{ij}\}}$, where $AM\{X_{ij}\}$ is the average value of export between provinces, while X_{ij} represents the true value of that flow and it is different for each pair of provinces, ij. Then, trade flows between each pair of provinces, ij, can be easily rewritten as:

$$X_{ij} = AM\{X_{ij}\}\gamma_{ij}. \tag{A.3}$$

Inserting Equation A.3 in Equation A.2:

$$GM\{X_{ij}\} = \left(\prod_{ij} AM\{X_{ij}\}\gamma_{ij}\right)^{\frac{1}{n_E n_I}} = AM\{X_{ij}\}\left(\prod_{ij}\gamma_{ij}\right)^{\frac{1}{n_E n_I}} = AM\{X_{ij}\} + GM\{\gamma_{ij}\}. \tag{A.4}$$

Rearranging Equation A.4 and taking the logs:

$$\log GM\{X_{ij}\} = \log\left[AM\{X_{ij}\} + GM(\gamma_{ij})\right] = \log AM\{X_{ij}\} + \log\left(1 + \frac{GM\{\gamma_{ij}\}}{A\{X_{ij}\}}\right) \tag{A.5}$$

Inserting Equation A.5 in Equation A.1:

$$\log AM\{X_{ij}\} = \frac{\sum_{i=1}^{n_E}\delta_i}{n_E} + \frac{\sum_{i=1}^{n_E}\sum_{j=1}^{n_I} O_l(\log Distance_{ij})}{n_E n_I} + \frac{\sum_{j=1}^{n_I}\delta_j}{n_I} + \frac{\sum_{i=1}^{n_E}\sum_{j=1}^{n_I} v_{ij}}{n_E n_I} \tag{A.6}$$

where:

$$\frac{\sum_{i=1}^{n_E}\sum_{j=1}^{n_I} v_{ij}}{n_E n_I} = \frac{\sum_{i=1}^{n_E}\sum_{j=1}^{n_I}\epsilon_{ij}}{n_E n_I} - \log\left(1 + \frac{GM\{\gamma_{ij}\}}{A\{X_{ij}\}}\right) \tag{A.7}$$

The resulting equation estimated for the first-step of the model takes the following form:

$$\log \overline{X}_{ij} = \frac{\sum_i \delta_i}{n_E} + \frac{\sum_{ij} O_l (\log Distance_{ij})}{n_E n_I} + \frac{\sum_j \delta_j}{n_I} + v_{ij}, \quad \text{(A.8)}$$

.

Appendix B. Total internal freight traffic by mode of transport

Table B1: Total Internal Freight Traffic by Mode of Transport, % Composition

Modes of Transport	2000	2005	2006	2007	2008	2009	2010
By Rail and Pipeline	16.11	14.18	15.68	15.96	15.63	13.19	13.02
By Road	**67.93**	**65.64**	**62.99**	**60.51**	**62.44**	**62.77**	**61.93**
Shipping	15.57	19.76	20.86	23.05	21.47	23.62	24.58
Air Transport	0,39	0.41	0.46	0.49	0.46	0.41	0.47
Total	100	100	100	100	100	100	100

Souce: Ministero delle Infrastrutture e dei Trasporti (2010-2011)

Appendix C. Total lenght of Roman road network by NUTS-3 Italian regions

Table C1: Lenght (km) of Roman road network

NUTS-3 Italian Region	*Major Roman Roads*	*All Roman Roads*	*NUTS-3 Italian Region*	*Major Roman Roads*	*All Roman Roads*
Agrigento	234.669	234.669	Medio Campidano	0	68.883
Alessandria	164.199	164.199	Messina	165.376	165.376
Ancona	0	58.611	Milano	71.773	147.889
Aosta	112.547	112.547	Modena	28.065	107.112
Arezzo	70.902	70.902	Monza e della Brianza	23.086	23.086
Ascoli Piceno	57.635	77.297	Napoli	81.127	128.403
Asti	58.857	58.857	Novara	13.121	51.588
Avellino	130.323	244.380	Nuoro	0	182.176
Bari	224.884	278.363	Ogliastra	0	85.579

Continued on next page

Table C1 continued from previous page

NUTS-3 Italian Region	*Major Roman Roads*	*All Roman Roads*	*NUTS-3 Italian Region*	*Major Roman Roads*	*All Roman Roads*
Barletta-Andria -Trani	73.196	136.123	Olbia-Tempio	0	379.264
Belluno	27.696	57.087	Oristano	0	249.726
Benevento	61.271	150.471	Padova	66.916	172.533
Bergamo	38.065	76.275	Palermo	263.677	263.677
Biella	7.108	7.108	Parma	35.253	101.363
Bologna	96.026	183.882	Pavia	137.353	192.269
Bolzano	266.310	266.310	Perugia	147.987	238.279
Brescia	95.769	228.561	Pesaro e Urbino	86.490	102.481
Brindisi	122.069	175.285	Pescara	40.232	53.961
Cagliari	0	255.505	Piacenza	92.579	165.950
Caltanissetta	97.815	97.815	Pisa	62.942	65.479
Campobasso	16.881	199.074	Pistoia	0	32.359
Carbonia-Iglesias	0	85.947	Pordenone	0	0
Caserta	193.996	355.327	Potenza	328.376	558.042
Catania	152.244	173.960	Prato	0.947	9.856
Catanzaro	137.174	174.057	Ragusa	118.623	161.176
Chieti	16.744	109.044	Ravenna	64.956	123.817
Como	85.739	99.482	Reggio di Calabria	184.132	249.846
Cosenza	313.013	528.684	Reggio nell'Emilia	33.422	62.504
Cremona	75.405	161.543	Rieti	102.729	127.275
Crotone	64.603	116.569	Rimini	39.180	48.626
Cuneo	200.106	200.106	Roma	604.252	2495.736
Enna	89.395	92.685	Rovigo	32.067	76.321
Fermo	0	26.700	Salerno	287.560	425.413
Ferrara	45.768	77.833	Sassari	0	260.658
Firenze	125.576	229.376	Savona	100.308	100.308
Foggia	264.029	567.897	Siena	18.146	18.146
Forlì-Cesena	50.431	69.390	Siracusa	149.385	175.365
Frosinone	86.590	385.984	Sondrio	53.220	56.392
Genova	104.916	104.916	Taranto	136.407	201.355
Gorizia	26.275	41.367	Teramo	50.843	134.082
Grosseto	96.562	109.892	Terni	110.788	152.770

Continued on next page

Table C1 continued from previous page

NUTS-3 Italian Region	*Major Roman Roads*	*All Roman Roads*	*NUTS-3 Italian Region*	*Major Roman Roads*	*All Roman Roads*
Imperia	34.477	34.477	Torino	254.221	366.377
Isernia	0	90	Trapani	181.557	187.406
La Spezia	64.185	65.454	Trento	138.160	138.160
L'Aquila	161.461	341.449	Treviso	134.704	198.328
Latina	158.943	670.213	Trieste	44.418	44.418
Lecce	79.709	148.393	Udine	203.388	271.206
Lecco	34.768	88.590	Varese	0	27.283
Livorno	111.531	111.531	Venezia	161.148	175.829
Lodi	78.058	99.773	Verbano-Cusio-Ossola	0	0
Lucca	25.288	47.578	Vercelli	73.668	91.443
Macerata	0	20.001	Verona	157.329	205.961
Mantova	50.519	116.392	Vibo Valentia	69.340	69.340
Massa-Carrara	15.994	54.064	Vicenza	35.777	51.527
Matera	76.923	259.500	Viterbo	133.183	271.421
			Total	10294.853	19505.315

Appendix D. Main variables used in Section 6.3

Table D1: Descriptive Statistics for the main variables used in Section "Persistence Channels"

VARIABLES	Obs.	Mean	sd	Source
Major Roman Roads (km)	110	94.59	91.74	Author's elaboration from DARMC
Roman Roads (km)	110	178.3	255.2	Author'selaboration from DARM
Highways (km)	110	64.98	61.57	Atlante Statistico Territorial ISTAT
Terrain Ruggedness Index	110	2.196	1.585	Autor's elaboration from Nunn and Puga (2012)
Longitude	110	12.01	2.676	ISTAT
Latitude	110	42.83	2.673	ISTAT
Surface (km^2)	110	2,746	1,589	ISTAT
Elevation Range (m)	110	742.1	497.0	ISTAT
Distance to Rome (km)	110	481.5	216.9	Author's computation with R "`osrm package`"
Social Capital Index	103	0.003	3.120	Cartocci (2007)
Institution Quality Index	103	0.587	0.213	Nifo and Vecchione (2014)
Number of extortion 2010	103	57.99	82.82	ISTAT
% Religious Wedding 2005	103	66.02	13.29	ISTAT
Urbanisation 1300 (Pop. in Thousand)	103	25.97	27.61	Malanima (2016)

Roads to Innovation: evidence from Italy[1]

1 Introduction

The role of transport infrastructure investments in fostering growth has been extensively studied in the economics and regional science literature. As documented by literature surveys, like Ferrari et al. (2019) and meta-analysis, like Melo et al. (2013) and Bom and Ligthart (2014), transport infrastructures have been found to display significant impacts on different economic outcomes. A reduction in transport costs associated to transport infrastructure investments can generate higher productivity of other inputs and lower production costs, can increase trade and competition by enlarging relevant markets and can favor the exploitation of scale economies. Moreover, greater accessibility contributes to raise the market potential of different locations, thus affecting the spatial allocation of human capital and economic activities (agglomeration economies); in particular, transport infrastructures can facilitate knowledge creation and diffusion. Indeed, more recent endogenous growth theory models rests on knowledge spillovers as one of the most important engine of growth (Romer, 1990; Aghion and Howitt, 1990; Acemoglu and Akcigit, 2012). Hence, it becomes crucial to understand if transport infrastructure investments, among other possible policy tools, are able to stimulate innovation and knowledge spillovers that are often constrained by geography (Jaffe et al., 1993). With few exceptions (e.g. Agrawal et al., 2017), this issue has been neglected by previous literature.

This study investigates the impact of road infrastructures (motorways) on the innovative capacity of Italian (NUTS-3) regions . This analysis requires to address a difficult identification issue linked to possible simultaneity between regional technological evolution and transport infrastructure investments. Indeed, such investments are typically not randomly allocated whenever governments tend to build infrastructures in low-income and low-innovation regions, or when high growth driven by local innovation fosters the demand for mobility and therefore the construction of highways. Moreover, there might be omitted factors that drive both infrastructure and innovation. In order to tackle this issue, we follow the historical route instrumental variable approach suggested by the urban and regional economics literature (Redding and Turner, 2015) and pioneered by Duranton and Turner (2012). This approach grounds on the idea that the presence of a transport network built in the past

[1] Joint work with Anna Bottasso, Maurizio Conti and Simone Robbiano

can be a good predictor for successive infrastructure investments[2].

In this study we consider the ancient Roman roads dating back to 117 A.D. as an instrument for the modern motorways endowment of the Italian regions[3]. Following the literature, we argue that the Roman road network is reasonably exogenous, given that Roman roads were built mainly for military purposes; therefore, conditionally on a set of geographic controls, we assume that there are not important local unobservables, that explain both the construction of Roman Roads in certain areas and their (very) long run patterns of growth. Our main result is that the stock of highways has a positive and significant impact on regional innovative activity. Estimates suggest that an increase of 10% in the length of the motorways network leads to a 1% increase in innovation as measured by regional patent fractional count[4]. Our main findings are confirmed when considering patents by region and technological fields as observation units and are robust to the inclusion of a set of geographic and inventor control variables. The impact of highways network is found to decline over time with the diffusion of information and communication technologies. Moreover, we find evidence of important heterogenous treatment effects, as we find that roads favor innovation particularly in fields characterized by a higher level of technological turnover and in regions where inventors are more scattered over the territory: this is exactly what we should find if we believe that roads foster innovation by making communication easier. Finally, results on possible displacement effects of transport infrastructure investments in one region are not conclusive, but seem to point in favor of the existence of negative spillovers on nearby regions' innovative activity.

This paper is organised as follow. Section 2 describes related literature, Section 3 presents our database and in Section 4 we describe the identification strategy. Results are presented in Section 5 which is followed by the conclusive section.

2 Related literature

This study is related to the literature on the effects of roads infrastructure on regional growth and productivity. The first study that properly addressed this issue is the one by Fernald (1999) on US data. His idea is that some industries rely more on road services so that they should be particularly affected by improvements in the road network. By applying this identification strategy he finds that regional productivity is positively affected by roads investments. Chandra and Thompson (2000) analyze US data at the county level and find that highways have a differential im-

[2]Studies that have followed this approach include Duranton et al. (2014), Duranton (2015), Agrawal et al. (2017), Baum-Snow et al. (2017) and Martincus et al. (2017).

[3]Such instruments has been firstly adopted by Garcia-López et al. (2015) when studying the impact of highways on the sub-urbanization of Spanish cities and have been successively employed in different studies like Percoco (2015), Holl (2016), Roca and Puga (2017), De Benedictis et al. (2018), Garcia-López (2019).

[4]Similar results are found by Agrawal et al. (2017) for US metropolitan statistical areas.

pact across industries and affect the spatial allocation of economic activity. They boost economic activity in the counties that they pass directly, although at the expenses of adjacent counties. Another important study conducted on US Metropolitan Statistical Area (MSA) by Duranton and Turner (2012) finds that an an increase in the stock of highways in the city leads to an increase in employment by about 1.5% after 20 years. Moreover, authors suggest that their result is unlikely to simply reflect the spatial reorganization of economic activity.

A recent paper by Ghani et al. (2016) evaluates the impact on productivity, employment, output and number of establishments of the so called Golden Quadrilateral (GQ) project, a recent major investment program which involved a massive upgrade of the GQ highways network. Authors finds that such investments significantly affected the growth of manufacturing activity. Another interesting study on the impact of the massive investments in highways in China is the one by Xu and Nakajima (2017) who find a positive effect of highways construction on investment and output, with notable differences across types of regions and industries.

Our study contributes to the literature that has analysed the impact of transport infrastructure on innovation. As already mentioned above, this issue has been analysed by very few works. In the seminal work by Agrawal et al. (2017) authors analyze the impact of interstate US highways on regional innovation. They apply an Instrumental Variable (IV) approach to overcome the endogeneity of highway endowment and find that 10% increase in interstate highways leads to a 1.7% increase in regional patenting activity over a five years period. In particular, they suggest that roads facilitate knowledge creation and diffusion also by favoring knowledge flows within metropolitan statistical areas. Following the spirit of Agrawal et al. (2017), Wang et al. (2018), using an IV approach, examine the impact of road development on innovation in China. In particular, to overcome endogeneity issues about road endowment, they use the mean slope in a city to measure the relative cost of road construction. Authors conclude that a 10% improvement in road density increases the average number of approved patents per company by 0.71%[5].

Recently some authors have investigated the relationship between innovation and rail infrastructure (e.g. Yamasaki, 2017; Dong et al., 2018; Lin, 2017). In particular, Yamasaki (2017) analyses the effect of rail access on the adoption of steam energy in relation to the expansion of the Japanese rail network between the late 1800s and early 1900s. Based on a Difference in Difference (*DiD*) strategy together with an IV approach to overcome the endogeneity of railway construction[6], authors suggest that

[5]Studies based on firm level data include Li et al. (2017) who suggest that roads have a positive impact on productivity in China by using an approach à la Fernald (1999) and Holl (2016), who evaluates the impact of roads on total factor productivity using Spanish data over a time span characterized by a significant expansion in the highways network. The author finds a negative and statistically significant effect of distance from highways on firm productivity.

[6]Authors construct their instrument by calculating the "cost-minimizing route" between destinations using slope information to account for costs of construction.

the growth of rail access from 1888 to 1892 accounts for 67% of the growth of steam energy from 1888 to 1902. Lin (2017) estimates a *DiD* model for a panel of Chinese cities observed over the period 2003-2013 in order to assess the impact of high-speed rail (HSR) on a number of economic outcomes, including patent applications, as a proxy for innovation activities within a city. Among other results, authors suggest that high-speed rail stimulates innovation by favoring greater scientific collaboration between cities and the diffusion of knowledge. Finally, in Dong et al. (2018), the relation between knowledge diffusion and the construction of China's high speed rail is assessed over the period 2006 - 2015. By instrumenting the construction of HSR with the spatial distribution of major military troop deployments in 2005 and Chinese railroad networks in 1962, authors show that in Chinese cities connected to the HSR network, researchers experienced a significant increase in productivity, in terms of quantity and quality of scientific publications.
Our study is the first one that analyze the impact of road infrastructure on regional innovation in Italy. Indeed, empirical evidence on this issue has never been provided other than on US data. In particular, we share the identification strategy based on historical route instrumental variable approach with Agrawal et al. (2017); however our observational units (NUTS 3 regions) are quite different from USA metropolitan statistical areas, from different perspectives, especially from a geographical point of view. Italian regions always share borders and are very heterogeneous in term of population density and economic development. Such characteristics allows us to better analyze the issue of spillover effects of transport investments on nearby regions.

3 Data

Our study analyze the relation between roads and innovation, as measured by patent fractional count per capita[7], on 95 Italian provinces (NUTS-3 regions) as defined in 1974[8]. The innovation literature recognizes patents as fundamental instruments of appropriation of the innovative activity; indeed, technologies with greater impact on welfare and economic development are more likely to be patented (Pakes and Griliches, 1980). However, patents measure inventions but do not measure all innovative activity (Smith, 2005) and not all inventions are patented. However, as argued by the innovation literature, patents are an effective measure of local technological capacity.
We recover annual data on patents from the European patent Office (EPO) database (EPO-Patstat) that includes bibliographical and legal status patent data on several

[7] The geographical distribution of patent applications is assigned according to the inventor place of residence. If more than one inventor characterizes a patent, the patent application is distributed equally between all of them and consequently between their provinces, eluding thus biased counting.

[8] As the number of Italian provinces has been progressively grown in recent years, we consider only the 1974 local Administrative setting.

countries at NUTS-3 regions level. Patent data refer to patent applications filed directly under the European Patent Convention or to patent applications filed under the Patent Co-operation Treaty and designating the EPO (Euro-PCT). A detailed set of information on applications, like the number of applicants and inventors and their characteristics, the relative technological IPC class of the patent[9] and NACE-2 statistical classification of economic activity are included. We recover patent data for the period 1978 - 2015 and we "regionalise" raw patent information by means of inventor address (NUTS3 codes). Data are finally categorised on the base of technological fields following the WIPO systematic technology classification, based on the codes of the International Patent Classification (IPC). In particular, we identify five patent classes, according to different technological fields, namely Electrical Engineering, Instruments, Chemistry, Mechanical Engineering and the residuals ones. Data are limited to 2015, since the two last years of available data underestimate application counts because of the delays in the publication of patent data (eighteen/twenty-four months since application)[10].
Turning to roads infrastructure regional endowment, we consider the total number of kilometres of motorways in each NUTS-3 region as provided by the Italian Central Institute of Statistics and the Automobile Club of Italy[11]. As regards data on the length of Roman roads, in particular on those defined as major/consular roads, we rely on the Digital Atlas of Roman and Medieval Civilization (DARMC), which provides georeferenced data at regional (NUT-3) level on the road network of the Roman Empire in 117 AD[12]. We then calculate the length of the major Roman roads in each Italian NUTS-3 region. Figure 1 shows the resulting map; in particular, in the Italian peninsula the total length of major roads is almost 10,000 kilometres.

Following Duranton and Turner (2012) and Agrawal et al. (2017), we include in the analysis a complete set of control variables, like (NUTS-3) regions surface, the difference between maximum and minimum altitude and an index of terrain ruggedness[13]. Table 1 shows the mean and the standard deviations of the main variables used in the study and Figure 2 depicts the territorial distribution of patents in 1988. It is worth noting that, since there are 6 provinces that do not have highways and for which the value of the major Roman roads is different from zero, our sample contains information on 89 provinces[14]. Indeed, we find that, even though some provinces have no motorways, they are nevertheless equipped with road infrastructure similar

[9] WIPO IPC-based technology field classification. Source: WIPO IPC Technology Concordance Table.

[10] See Bronzini and Piselli (2016) for more details.

[11] https://ebiblio.istat.it/SebinaOpac/resource/statistica-degli-incidenti-stradali/IST0010868

[12] The main predecessor of this database is Talbert (2000) which provides maps of the entire Greek and Roman empires, covering the territory of over 75 modern countries.

[13] Authors' elaboration from Nunn and Puga (2012).

[14] In particular, we drop Brindisi, Matera, Agrigento, Ragusa, Siracusa and Grosseto from the sample.

Figure 1: Roman Road Network in Italy: Major Roads.

Source: Authors' elaboration from McCormick et al. (2013)

to them. These infrastructures are however not easily measurable, leading to potentially biased estimates.

Figure 2: NUTS-3 Patent Fractional Count (per million people) 1988

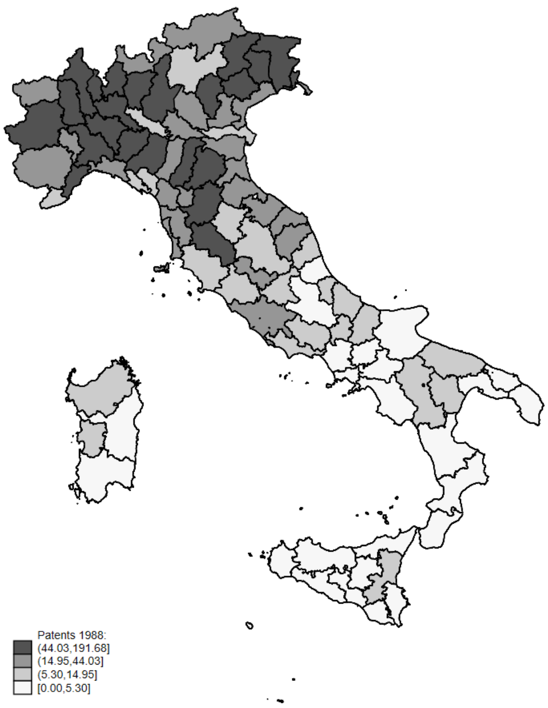


Table 1: Summary Statistics for Main Variables

Variables	Obs.	Mean	SD
Patent Fractional Count (per million people) 1983	89	14.62	14.33
Patent Fractional Count (per million people) 1988	89	31.36	35.91
Number of Inventors (per million people) 1988	89	21.01	24.95
Motorways lenght (km)	89	67.30	53.96
Major Roman Roads (km)	89	108.3	98.39
Surface (km^2)	89	2,931	1,594
Range (m)	89	758.8	510.6
Terrain Rugedness Index	89	2.299	1.605

Notes: The final number of observations we use for our computations is 89, since there are 6 provinces that do not have highways and for which the value of the major Roman roads is different from zero. Indeed, analysing these provinces, we find that, even though they have no motorways, they are nevertheless equipped with road infrastructure similar to them. These infrastructures are however not easily measurable, leading to potentially biased estimates. For these reasons, we drop Brindisi, Matera, Agrigento, Ragusa, Siracusa and Grosseto from the sample.

4 Identification Strategy

Following Agrawal et al. (2017), we consider an innovation model[15] where the innovative activity in each Italian province at time t is linked to the length of the motorways' system at time $t-j$:

$$\log Innov_{p,t} = \alpha + \beta\big(Motorways_{p,t-j}\big) + \gamma\big(Innov_{p,t-j}\big) + \varphi X_p + v_p \quad (1)$$

In Eq. (1), $Innov_{p,t}$ refers to the natural logarithm of our innovative measure in province p for patents at time t, while $Motorways_{p,t-j}$ is the length (logarithm) of the motorways' endowment for province p at time $t-j$. Following a common practice in innovation literature, also the lagged dependent variable is included as an input for future knowledge, in order to take into account the cumulative nature of the latter[16]. Moreover, the lagged dependent variable should account for most of the time invariant unobserved heterogeneity at the province level[17]. The econometric model includes a set of inventor and geography control variables, X_p. In particular, inventor controls include a measure for the number of inventors residing in each province, while geography controls include measures for surface, terrain asperity and elevation.

The β coefficient is our parameter of interest and refers to the impact of motorways endowment on local innovative activity: precisely, β measures the rate at which local innovative activity responds to motorways provision. Therefore, we interpret β as the parameter associating the endowment of motorways at time t with innovation growth for the period $t+j$.

Eq. (1) assumes that the analysis is conducted at an aggregate level, i.e. considering the sum of the patents originating in a given province, without variability between sectors. To allow an indication of the relevance of roads for knowledge in a given sector, we also estimate the following equation:

$$\log Innov_{p,f,t} = \alpha + \beta\big(Motorways_{p,t-j}\big) + \gamma\big(Innov_{p,f,t-j}\big) + \varphi X_{p,f} + v_{p,f} \quad (2)$$

where $Innov_{p,f,t}$ refers to the natural logarithm of our innovative measure in province p for patents in technological field f at time t.

As Drucker (2016) suggests, the effects of innovation activities, as R&D or public research, may spread over space and that spillovers are substantial up to 97 kilometres. For this reason, it is possible that innovation can be also displaced from one province to another. Indeed, the provision of motorways can potentially generate a zero-sum game among provinces. Therefore, in the spirit of Moretti and Wilson

[15] We provide an in depth explanation of the model in Appendix A.

[16] See Aghion and Howitt (1990).

[17] Since the diffusion of knowledge is supported by face-to-face interactions favoured by the presence of highways, one might also expect to find a measure of market potential. Nevertheless, we believe that the presence of the lagged dependent variable can control for this aspect.

(2014) and Agrawal et al. (2017) we extend our main model in Eq. (1) with the inclusion of spatial lag in order to analyze if the impact of roads infrastructure investments generate spillover effects in nearby regions. Specifically, we estimate the following model[18]:

$$\log Innov_{p,t} = \alpha + \beta(Motorways_{p,t-j}) + \gamma(Innov_{p,t-j}) + \\ + \theta SpatialMotorways_{p,t-j} + \varphi X_p + v_p \tag{3}$$

where:

$$SpatialMotorways_{p,t-j} = \sum_{p \neq i}^{P} w_{pi} \log Motorways_{i,t-j} \tag{4}$$

The additional term $SpatialMotorways_{p,t-j}$ represents, for each province p, a weighted average of the motorways stock in other provinces i at time t; it is based on a row normalized matrix of the inverse of the distances multiplied by the levels of innovative capacity between any province p and i, with elements $w_{pi} = (Patent_p * Patent_i)/Distance_{pi}$ such that $\sum_{p \neq i}^{P} w_{pi} = 1$. Notice that to address possible endogenity issues we consider the innovative capacity at the begginning of the sample period, as suggested by Corrado and Fingleton (2012) and Bottasso et al. (2014). Thus, the spatial lag of motorways endowment account for the the possibility that infrastructural endowment in a certain province might have an impact on the performance of nearby provinces. The θ coefficient associated to the spatial lag allows us to detect the nature of possible spillover effects. A negative and significant sign of this parameter would suggest the presence of significant spatial displacement effects on regional innovation generated by road transport infrastructure investments in a particular region.

Estimating the effect of road transport infrastructure investments on regional innovative capacity is a quite challenging task in terms of identification strategy. Indeed, there might be simultaneity between regional technological evolution and transport infrastructure investments. In fact, such investments are typically not randomly allocated whenever governments tend to invest in lagging areas or in low-income and low-innovation regions, or when high growth driven by local innovation fosters the demand for mobility and therefore the construction of highways. Moreover, there might be omitted factors that drive both infrastructure and innovation. Thus, in Eq. (1) and (3) there might be correlation between unobservables, $v_{p,f}$, and the endowment of motorways in a province. In this case, estimating Eq. (1) and (3) by OLS could generate an underestimation of the causal impact of motorways on innovative performances. To take into account such a concern, we implement one of three approaches usually adopted in the applied literature and described by Redding and Turner (2015), namely, the historical route instrumental variables, the planned

[18] As for Eq. (1), also the Eq. (3) is estimated not only at an aggregate level but also allowing variability between technological fields: $\log Innov_{p,f,t} = \alpha + \beta(Motorways_{p,t-j}) + \gamma(Innov_{p,f,t-j}) + \theta SpatialMotorways_{p,t-j} + \varphi X_{p,f} + v_{p,f}$.

route instrumental variables and the inconsequential units approach.

In this work we rely on the first approach and we use the ancient Roman roads length dating back to 117 AD as an instrument for the current motorways endowment of the Italian provinces. Moreover, for each province p, we build the following instrument for the spatial lag presented in Eq. (4):

$$SpatialRomanRoads_{p,t} = \sum_{p \neq i}^{P} w_{pi} \log RomanRoads_{i,t}, \tag{5}$$

which is a weighted average for Roman roads endowment in other NUTS-3 regions[19]. This approach has been pioneered by Duranton and Turner (2011, 2012) that choose the routes of major expeditions of exploration between 1835 and 1850 and the major rail routes in 1898 as instruments for MSA highways endowment. These instruments have also been used in subsequent works, e.g. Duranton et al. (2014), Duranton (2015) and Agrawal et al. (2017).

Other ancient transport network measures have been proposed by the literature as instruments for current roads endowment. Baum-Snow et al. (2017) analyze how urban railroads and highways have influenced urban form in Chinese cities by using the 1962 Chinese transport network as instrument, while Martincus et al. (2017) consider the Inca roads built before 1530 as an instrument for the 2000s Peruvian transport infrastructure.

Among the various instruments used in the literature, one in particular has seen numerous applications, i.e. the ancient Roman road network. One of the first uses of the ancient Roman network as an instrument for the current transport system is represented by Garcia-López et al. (2015). In their work, the authors investigate the effect of highways on the suburbanization of Spanish cities by relying on an IV approach where the instrument is represented by Spain historical roads, namely the old Roman roads and the roads built by the Bourbons in the XVIII century[20]. Subsequent works that have used the roman road network within an IV approach include Percoco (2015), Holl (2016), Roca and Puga (2017), De Benedictis et al. (2018), and Garcia-López (2019).

The validity of the Roman roads network as an instrument for modern roads endowment has been largely discussed by the aforementioned studies, both in terms of relevance and exogeneity. Regarding the first characteristic, i.e. the relevance, as common sense suggests, historical transport networks might be relevant because modern networks are not built in isolation from them (Garcia-López, 2019). This hypothesis has been tested by various studies. In particular a positive correlation between Roman roads and current Spanish highways has been shown by Garcia-López

[19]The weighting matrix has been constructed as the spatial matrix defined in Eq.(4) using Roman roads instead of modern roads.

[20]For a previous application see Garcia-López (2012), where the author limit the analysis to the metropolitan area of Barcelona.

et al. (2015); Garcia-López (2019); Holl (2016); Holl and Mariotti (2018). Also Percoco (2015) and De Benedictis et al. (2018) have found a strong relationship between current and Roman roads network in Italy[21]. We confirm these findings, as suggested by first stage regressions diagnostic. The first-stage F-statistics confirm that weak-identification bias is not a problem, since their values overall exceed thresholds proposed by Stock and Yogo (2005).
Another condition that our instrument has to satisfy is the exclusion restriction, whereby it should affect regional innovation only through its effect on the current highway endowment; moreover, it should be independent from contemporaneous level of innovation activity at the NUTS-3 regional level. Indeed, the validity of the instrument requires its exogeneity conditional on controls and, according to previous literature, this requirement seems to be satisfied by our chosen instrument. In particular, as argued by Dalgaard et al. (2018), Roman roads are strongly predetermined and, more in general, almost any ancient transport network may be considered as exogenous because of the time that has elapsed since it was built (Duranton and Turner, 2012). Moreover, the literature has identified military reasons as the main purposes of Roman road construction, thus excluding a direct economic reason for their location (e.g. Garcia-López et al., 2015; De Benedictis et al., 2018). However, since geography may have influenced the construction of both Roman roads and modern motorways, the exogeneity of our instruments is also based on the inclusion of a set of geographical controls, as in De Benedictis et al. (2018) and Garcia-López (2019).

5 Empirical Results

5.1 Main Results

Following Agrawal et al. (2017), among other authors, we estimate our model for the year 1988, which precedes the large diffusion of the Information and Communication Technology (ICT) and we use a five year lag both for the motorways stock and the lagged dependent variable. Indeed, the ICT revolution might have boosted or reduced the impact of road infrastructures on innovative activity, depending on the substitutability or the complementarity between personal interactions and ICT in the knowledge production process.
We first estimate Equations (1) and (3) relying on a measure of patent fractional count aggregated at NUTS-3 regional level, without considering the technological field of each patent. In column (1) of Table 2 we show OLS estimates that do not include controls, while in column (2) we report results on estimates that include ge-

[21]Percoco (2015) uses Roman roads as an instrument for road accessibility, as measured by the presence of a motorway exit, while De Benedictis et al. (2018) use Roman roads as an instrument for modern roads endowment.

ographic and inventors controls. Parameter estimates show evidence in favor of a positive correlation between 1983 motorways stock and the patent fractional count in 1988. In column (3) and (4) we present estimates of Eq. (3). The inclusion of the spatial lag, does not affect the positive correlation between highways and innovation; the coefficient of $logMotorways_{p,1983}$ remains positive and significant, while the coefficient of $SpatialMotorways_{p,1983}$ is not significant.

Table 2: Main Results for the Impact of Motorways on the Inovative Activity: OLS Estimation.

	(1) OLS	(2) OLS	(3) OLS	(4) OLS
Dependent Variable: log Patents Fractional Count (per capita) 1988				
$logMotorways_{p,1983}$	0.0930**	0.0967**	0.0896**	0.0928**
	(0.0387)	(0.0396)	(0.0388)	(0.0397)
$SpatialMotorways_{p,1983}$			0.410	0.396
			(0.397)	(0.413)
Geography	NO	YES	NO	YES
Inventors	NO	YES	NO	YES
Observations	89	89	89	89
R-squared	0.777	0.783	0.779	0.784

Notes: All specifications include the lagged dependent variable. Inventor controls include the log of the inventors (per capita) in each NUTS-3 region in 1983. Geography controls include surface, terrain ruggedness and elevation. We add one to all patent, inventor and motorways before to taking the log to include observations with value of zero. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Results in Table 2 show a positive relation between road infrastructure and regional innovative activity; however such results might be biased because of endogeneity issues discussed above. In order to interpret our results in a causal way we need to rely on instrumental variable estimates.

Table 3 presents IV results and shows that the coefficient of $logMotorways_{p,1983}$ is not different from zero in column (1), while it is stable and significant across all other specifications, ranging from 0.10 to 0.13 values.

These results suggest that the endowment of highways has a positive and significant impact on regional innovative activity. In particular, estimates imply that an increase of 10% in the length of the motorways network in 1983 leads to an increase of about 1% in 1988 regional patent fractional count. Therefore, neither the inclusion of control variables nor the spatial lag influence the positive impact of highways. Moreover, the coefficient of the spatial lag variable ($SpatialMotorways_{p,1983}$) is not statistically different from zero, thus suggesting the absence of significant spatial displacement effects.

It is worth noting that, following Agrawal et al. (2017), the analysis focuses mainly on innovative capacity in 1988, but it is only since the early 1990s that Italy has

Table 3: Main Results for the Impact of Motorways on the Inovative Activity: IV Estimation.

	(1)	(2)	(3)	(4)
	IV	IV	IV	IV
Dependent Variable: log Patents Fractional Count (per capita) 1988				
$logMotorways_{p,1983}$	0.0873	0.102*	0.106*	0.130**
	(0.0534)	(0.0582)	(0.0558)	(0.0619)
$SpatialMotorways_{p,1983}$			-0.627	-0.742
			(0.461)	(0.497)
Geography	NO	YES	NO	YES
Inventors	NO	YES	NO	YES
Observations	89	89	89	89
R-squared	0.777	0.782	0.767	0.771
F-statistic	30.45	24.28	10.95	8.549

Notes: All specifications include the lagged dependent variable. Inventor controls include the log of the inventors (per capita) in each NUTS-3 region in 1983. Geography controls include surface, terrain ruggedness and elevation. We add one to all patent, inventor and motorways before to taking the log to include observations with value of zero. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

increased its percentage share of applications to the EPO by reducing its distance from other industrialized nations (Larédo and Mustar, 2001). Therefore, it is reasonable to think that our estimates may suffer from a substantial underestimation of the innovative activity of small and medium enterprises, which represent the largest share of Italian companies and which are the ones that could benefit most from the expansion of roads.

In order to estimate Eq. (1), we consider one excluded instrument, namely the (log) length of ancient Roman roads in each province, to account for the possible endogeneity of current highways. Moreover, estimation of Eq. (3) (columns 3 and 4) requires an additional excluded instrument, i.e. $SpatialRomanRoads_p$ as defined in Eq. (5). Chosen instruments result to be relevant as the first stage F-statistics in Table 3 suggest. Indeed, the F-statistic ("Kleibergen-Paap rk Wald F statistic") exceeds all critical values for the weak instrument test based on TSLS size (Stock and Yogo, 2005).

It is worth noting that our results are also robust to the adoption of a Limited Information Maximum Likelihood (LIML) estimation method. In addition, the choice to "regionalise" patent information via the inventor address could lead to misleading estimates. In fact, the presence of a specific motorway may increase the patent activity observed in a specific province, p. However, some of the inventors living in that province may work in a neighbouring province, i. So the real impact of the motorway is to allow inventors to live in a given province, p, while working in the neighbouring one, i. By construction, in some cases there may be an increase in innovative activity in the province p when it is attributable to the province i. Therefore, to verify the robustness of our results we replicate the analysis using another criterion to construct the dependent variable, i.e. we use the address of the applicants. The main results are confirmed, as shown in the Table B1 in Appendix B.

To reduce possible unobserved heterogeneity and to reinforce the exclusion restriction assumption, we turn to analyze the relation between the highways stock and patent fractional count in each province p, at time t, in technological field f, according to Eq. (2).

Table 4: Results for the Impact of Motorways on the Innovative Activity by Province and Technological field: IV Estimation.

	(1) IV	(2) IV	(3) IV	(4) IV
Dependent Variable: log Patents Fractional Count (per capita) 1988 by Province and Tech. Field				
$logMotorways_{p,1983}$	0.0536	0.0759**	0.0879*	0.107**
	(0.0429)	(0.0366)	(0.0513)	(0.0425)
$SpatialMotorways_{p,1983}$			-0.993*	-0.744**
			(0.509)	(0.368)
Field FE	YES	YES	YES	YES
Geography	NO	YES	NO	YES
Inventors	NO	YES	NO	YES
Observations	356	356	356	356
R-squared	0.634	0.690	0.595	0.676
F-statistic	19.26	16.90	8.815	7.138

Notes: All specifications include the lagged dependent variable. Field FE refers to four technological fields classes. Inventor controls include the log of the inventors (per capita) in field f in each NUTS-3 region in 1983 and the total (log) number of inventors in each NUTS-3 region in 1983. Geography controls include surface, terrain ruggedness and elevation. We add one to all patent, inventor and motorways before to taking the log to include observations with value of zero. Robust standard errors clustered at the NUTS-3 region level in parentheses *** p<0.01, ** p<0.05, * p<0.1

We consider four different technological classes, constructed by following WIPO systematic technology classification, namely Electrical Engineering, Instruments, Chemistry, Mechanical Engineering[22]. In Table 4 we report results obtained after replicating previous analysis on observation units defined at province and technological field level. Estimated models include also technological field fixed effects and a specific control for the number of inventors in each field[23].

All estimated specifications confirm previous findings and suggest the existence of a positive and significant effect of motorways endowment on innovation activity. The coefficients of $logMotorways_{p,1983}$ are slightly smaller with respect to those reported in Table 3. Moreover, the coefficients of the spatial lag variable are negative and significant, thus suggesting the existence of negative spillovers, whereby one province motorways endowment negatively affects innovation activity in nearby provinces.

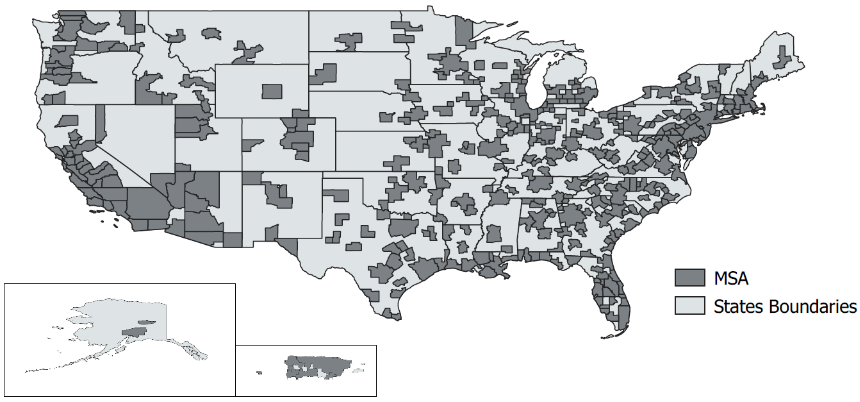
[22]The database is obtained by disaggregating patent fractional count on the basis of the technological class. It is worth noting that not all provinces in 1988 patented in all fields.

[23]As a consequence, inventors control variables include both the total number of inventors in each province and the number of inventors in each province in specific field f.

These results can be interpreted as evidence in favor of spatial reorganization of economic activities associated to road transport investments that generate zero-sum game among provinces. Such evidence on displacement effects is indeed not conclusive, given that our estimates based on observation units not taking into account patent technological field do not provide the same result.

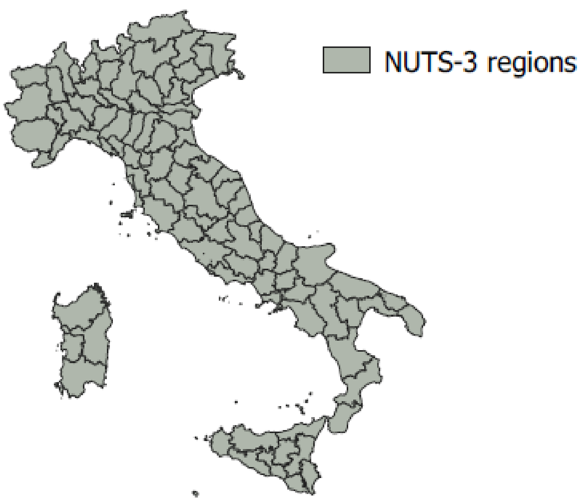
Our (not conclusive) results on the existence of spatial spillovers differ from Agrawal et al. (2017), where no significant displacement effects are detected. A plausible explanation for this discrepancy could lie in the different characteristics of US MSA and Italian NUTS-3 regions.

Figure 3: Metropolitan Statistical Area in the U.S.



Source: Authors' elaboration from US Census Bureau

Figure 4: NUTS-3 Regions in Italy



Source: Authors' elaboration from ISTAT

Indeed, as shown in Figure 3, the US MSA are more sparse across the country and not all of them share borders. In this context, the presence of road infrastructures facilitate ideas' circulation within the same MSA, but does not affect the probability that those ideas might be used by someone else in a different MSA. On the other side, the geography of Italian provinces (Figure 4) significantly differs from that of MSA. Indeed, this administrative division covers the whole national territory and each province shares at least one border with another province. In this context, it seems more plausible that local innovation gains from motorways endowment in one region might be offset by losses in nearby regions. Nevertheless, our results have to be interpreted with caution since in this analysis we primarily assess the local impact of highways rather than the national impact of regional motorways. Indeed, as argued by Agrawal et al. (2017), conducing a precise evaluation of the effect of highways at national level is very challenging and would require a general equilibrium type of approach.

5.2 Heterogeneous Effects

In the spirit of Agrawal et al. (2017), we explore possible heterogeneous effects of roads on innovation. We first investigate whether roads differently affect innovation in industries characterized by a faster technological turnover. Since time is needed for additional knowledge to spread sufficiently to be useful for other innovators (Caballero and Jaffe, 1993), motorways should play an even greater role for knowledge diffusion in industries characterized by a faster process of *"creative destruction"*. In these industries the very high speed of technological turnover tends to limit the usefulness of knowledge flows, unless ideas can easily spread; in these contexts, the road network might potentially mitigate this problem by facilitating faster knowledge flows.

In the seminal work by Hall et al. (2001), authors measure the obsolescence of knowledge by technology field, finding that the speed of technological turnover varies across industries. In particular, according to the categories they developed, Computer & Communications and Electrical & Electronics industries show the highest technological turnover rate, while Drugs & Medical, Chemical and Mechanical industries present lower rates. In this spirit, we assume Electrical Engineering to be an "High-Tech Turnover" industry, while Instruments, Chemistry and Mechanical Engineering to be "Low-Tech Turnover" industries. We then split the sample according to this classification and we estimate Eq. (2) for each sub-sample. Results from IV estimation are presented in Table 5.

Estimates reported in column (1) show that a 10% increase in the motorways endowment in 1983 leads to an increase of 1% in patent fractional count in 1988 for industries where technological turnover is high. On the contrary, the coefficient of $logMotorways_{p,1983}$ in column (2) is not different from zero, thus suggesting that 1983 motorways does not affect local innovation levels in Low-Tech Turnover industries. This result is line with Agrawal et al. (2017) obtained don US MSA[24].

[24] As shown in the Appendix A, Agrawal et al. (2017) empirical framework relies on a model in which regional innovative activity is related to the level of highways through the relationship $K_t^* =$

Table 5: Results for the Heterogenous Impact of Motorways on the Innovative Activity at Technological Field Level: IV estimates High/Low Velocity Industries.

	(1)	(2)
Sample	High-Tech Turnover Industries	Low-Tech Turnover Industries
	IV	IV
Dependent Variable: log Patents Fractional Count (per capita) 1988 by Tech. Field		
$logMotorways_{p,1983}$	0.104*	0.0604
	(0.0629)	(0.0457)
Field FE		YES
Geography	YES	YES
Inventors	YES	YES
Observations	89	267
R-squared	0.526	0.719
F-statistic	12.96	17.29

All specifications include the lagged dependent variable. High velocity patent field refers only to Electrical Engeneering and Communication technology while low velocity ones include Instruments, Chemistry and Mechanical Engeneering technologies. Field FE are then computed only for low velocity subsample. Inventor controls include the log of the inventors (per capita) in field f in each NUTS-3 region in 1983 and the total (log) number of inventors in each NUTS-3 region in 1983. Geography controls include surface, terrain ruggedness and elevation. We add one to all patents, inventors and motorways before to taking the log to include observations with value of zero. Robust standard errors clustered at the NUTS-3 region level in parentheses *** p<0.01, ** p<0.05, * p<0.1

We further investigate weather the impact of motorways on innovation is heterogeneous according to different degrees of inventors density. Indeed, geographic proximity favors the development of knowledge flows, learning processes and relations between inventors, which in turn affect innovative activity. In this context, motorways system represents an important tool in facilitating the creation of networks between scientist and organisations.

We construct a measure of inventor density, $(\frac{Inventors_{p,1983}}{Surface_p})$, to capture the degree of inventors dispersion over the regional territory and we split the sample in High-Density or Low-Density regions depending on whether they are above or below the mean inventor density. We expect that the provision of highways would benefits more regions where interactions between inventors require large travelling distances.

Results, reported in Table 6, are in line with Agrawal et al. (2017) and show a zero impact of motorways in High-Density regions, while for Low-Density provinces a 10% increase in motorways endowment leads to a roughly 1.2% increase in patent fractional count.

$A * Motorways_t^{\alpha}$. In such a framework, authors demonstrate that the relation between the coefficient of $logMotorways_{p,t}$, β and the coefficient of the lagged dependent variable, γ, is $\beta = \alpha(1-\gamma)$. Since High-Tech Turnover industries are expected to exhibit lower values of γ, the β coefficient is expected to be larger.

Table 6: Results for the Heterogeneous Impact of Motorways on the Innovative Activity: IV estimates High/Low Density NUTS-3 Regions.

Sample	(1) High Density Regions IV	(2) Low Density Regions IV
Dependent Variable: log Patents Fractional Count (per capita) 1988		
$logMotorways_{p,1983}$	-0.0340	0.117**
	(0.122)	(0.0584)
Geography	YES	YES
Inventors	YES	YES
Observations	18	71
R-squared	0.901	0.686
Geography	YES	YES
Inventors	YES	YES
F-statistic	4.860	27.48

All specifications include the lagged dependent variable. Low density NUTS-3 regions are the ones where the inventor density ($\frac{Inventors_{p,1983}}{Surface_p}$) is below the average value of the whole sample. Respectively, high density NUTS-3 regions show values above the average values. Inventor controls include the log of the inventors (per capita) in field f in each NUTS-3 region in 1983 and the total (log) number of inventors in each NUTS-3 region in 1983. Geography controls include surface, terrain ruggedness and elevation. We add one to all patent, inventor and motorways before to taking the log to include observations with value of zero. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

5.3 ICT and Roads

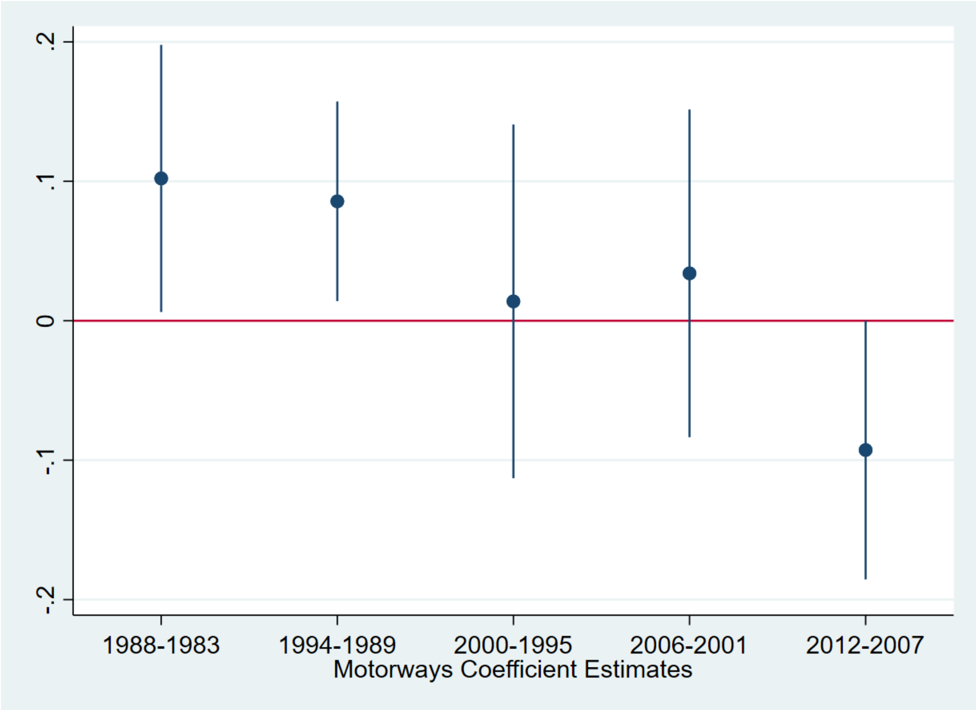
As already mentioned, in this study we focus on the relationship between regional innovative capacity observed in 1988 and the 1983 regional motorways stock. This choice accounts for the fact that, in these years, knowledge flows are supposed not to be affected by the Internet revolution of the 1990s. In fact, our study relies on the idea that roads have a prominent role in favouring knowledge flows, which in turns promote innovative capacity. Since the ICT has brought about revolutionary changes in the way people work, communicate, learn, spend time and interact (Jorgenson and Vu, 2016), it is interesting to investigate whether road infrastructures still have an important role in shaping knowledge flows when new communication technologies have been made available.

We turn to investigate the long term impact of highways, once the ICT revolution occured, by exploring the impact of 1983 regional highways endowment on innovation levels in different subsequent years. We estimate the following model for five different time periods (1989-1983, 1994-1989, 2000-1995, 2006-2001 and 2012-2007):

$$\log Innov_{p,t} = \alpha + \beta\big(Motorways_{p,1983}\big) + \gamma\big(Innov_{p,t-j}\big) + \varphi X_p + \upsilon_p \quad (6)$$

Figure 5 shows IV estimates for the $Motorways_{p,1983}$ coefficient over five differ-

Figure 5: Effect of 1983 Motorways endowment on level of innovation across time



Notes: the figure displays IV estimates for the motorways coefficient in five periods: 1988-1983, 1994-1989, 2000-1995, 2006-2001 and 2012-2007. The equation computed for each period is: $\log Innov_{p,t} = \alpha + \beta Motorways_{p,1983} + \gamma Innov_{p,t-5} + \varphi X_p + \upsilon_p$. Confidence intervals are reported at 90% level.

ent years. Compared to 1988, in 1994 the coefficient slightly decreases in magnitude, but remains positive and significant. This reflects the fact that the Internet revolution is still in its early phases, especially in Italy. The transition to digital media appears not yet complete and, therefore, the role of road infrastructure remains persistent. On the other hand, ten years after the ICT revolution the role of 1983 motorways have no longer effects on the levels of innovation in 2000. The same result is also evident for the year 2006. It is worth noting that the coefficient of $Motorways_{p,1983}$ for year 2012 is negative. The interpretation of this result, however, is very challenging and should deserve further investigations.

In this context, we interpret such result considering that, in the last twenty years, the focal issue about transport systems is changed from a quantitative perspective to a qualitative one. It is well-known that many Italian provinces face problems of traffic congestion and the presence of highways is not always indicative of adequate mobility. Furthermore, there is increasing discussion about the role of quality of life in the locational choices of activities with a high scientific and technological content. Consequently, given the advent of ICT, which allows instant communications around the world, the choice of innovators on where to work and live is increasingly based on motivations related to sustainability and quality of life (Knox and Mayer, 2013), rather than those purely related to the possibility of using road infrastructure. This trend could open up new opportunities for urban centres that make the combination of high technology and liveability a new model of urban development[25].

6 Conclusions

In this work we assess the impact of motorways endowment on innovative capacity. We explore such relation by estimating a model that links the local innovative capacity, proxied by patent fractional count, at time t with the stock of road infrastructures at time $t-j$. We conduct our analysis on a cross-section of Italian NUTS-3 regions. In particular, our main specification analyzes the relationship between regional innovative capacity in 1988 and regional road stock in 1983. We include in the model the lagged dependent variable and we control for the number of inventors and for a set of geographic charachateristics. We first conduct our analysis with an aggregate measure of patent fractional count in each province, with no distinction on patents technological fields. Secondly, to allow for a finer analysis, we account for technological fields categorisation. The main challenging issue about the estimation of our models arises from the possible endogeneity of highways stock. To deal with this problem, we follow the "historical instrumental variable" approach by using the length of the ancient Roman roads system dating back to 117 AD as an instrument for the length of current motorways. The chosen instrument results to be relevant as our first stage regressions suggest, thus confirming that modern network are not built in isolation from old roads network. Moreover, the validity of such instrument has been largely discussed in the previous literature. Instrumental variables estimates of our baseline specification indicates that 1983 highways network has a positive and significant impact on 1988 innovative capacity, in particular an increase of 10% in the length of motorways corresponds to a 1% increase in patent fractional count.

We further analyse the possible existence of spatial spillovers by introducing in our model a "spatial lag" that accounts for motorways stock in nearby provinces. Results show mixed evidence for potential displacement effects.

We further investigate whether the impact of highways on innovation shows some heterogeneity across provinces and industries. Estimates conducted on different sample splits suggest that motorways endowment has a positive effect on industries where the technological turnover is faster, while no effects are found for industries characterized by a lower technological turnover. Moreover, we find that highways benefit more inventors located in regions where local interaction requires long distances to be travelled.

[25]However, it is possible that results in the 2000s are biased because of the non-negligible expansion of the highways network that has characterized certain areas of the country in the 1990s: if highways construction has been relatively more important in regions characterized by a low density network in the 1980s, the coefficient of the 1983 highways network might be biased. We plan to test for this in an extension of the current paper.

Finally, we investigate the long term impact of 1983 motorways on 1994, 2000, 2006 and 2012 innovative capacity, finding that, due to the ICT revolution of the 1990s, motorways do not seems to display any causal effect on recent innovation outcomes.

References

Acemoglu, D. and Akcigit, U. (2012), 'Intellectual property rights policy, competition and innovation', *Journal of the European Economic Association* **10**(1), 1–42.

Aghion, P. and Howitt, P. (1990), A model of growth through creative destruction, Technical report, National Bureau of Economic Research.

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

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

Bom, P. R. and Ligthart, J. E. (2014), 'What have we learned from three decades of research on the productivity of public capital?', *Journal of economic surveys* **28**(5), 889–916.

Bottasso, A., Conti, M., Ferrari, C. and Tei, A. (2014), 'Ports and regional development: a spatial analysis on a panel of european regions', *Transportation Research Part A: Policy and Practice* **65**, 44–55.

Bronzini, R. and Piselli, P. (2016), 'The impact of r&d subsidies on firm innovation', *Research Policy* **45**(2), 442–457.

Caballero, R. J. and Jaffe, A. B. (1993), 'How high are the giants' shoulders: An empirical assessment of knowledge spillovers and creative destruction in a model of economic growth', *NBER macroeconomics annual* **8**, 15–74.

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

Corrado, L. and Fingleton, B. (2012), 'Where is the economics in spatial econometrics?', *Journal of Regional Science* **52**(2), 210–239.

Dalgaard, C.-J., Kaarsen, N., Olsson, O. and Selaya, P. (2018), Roman Roads to Prosperity: Persistence and Non-Persistence of Public Goods Provision, CEPR Discussion Papers 12745, C.E.P.R.
URL: *https://ideas.repec.org/p/cpr/ceprdp/12745.html*

De Benedictis, L., Licio, V., Pinna, A. M. et al. (2018), The long-term effects of the historical roman road network: trade costs of italian provinces, Working Paper CRENoS 201801, Centre for North South Economic Research, University of Cagliari and Sassari.

Dong, X., Zheng, S. and Kahn, M. E. (2018), The role of transportation speed in facilitating high skilled teamwork, Technical report, National Bureau of Economic Research.

Drucker, J. (2016), 'Reconsidering the regional economic development impacts of higher education institutions in the united states', *Regional Studies* **50**(7), 1185–1202.

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

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

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

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

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

Ferrari, C., Bottasso, A., Conti, M. and Tei, A. (2019), *Economic Role of Transport Infrastructure: Theory and Models*, Elsevier, Amsterdam.

Garcia-López, M.-A. (2012), 'Urban spatial structure, suburbanization and transportation in barcelona', *Journal of Urban Economics* **72**(2-3), 176–190.

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. and 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. and 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.

Hall, B. H., Jaffe, A. B. and Trajtenberg, M. (2001), The nber patent citation data file: Lessons, insights and methodological tools, Technical report, National Bureau of Economic Research.

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

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

Jaffe, A. B., Trajtenberg, M. and Henderson, R. (1993), 'Geographic localization of knowledge spillovers as evidenced by patent citations', *the Quarterly journal of Economics* **108**(3), 577–598.

Jorgenson, D. W. and Vu, K. M. (2016), 'The ict revolution, world economic growth, and policy issues', *Telecommunications Policy* **40**(5), 383–397.

Knox, P. and Mayer, H. (2013), *Small town sustainability: Economic, social, and environmental innovation*, Walter de Gruyter.

Larédo, P. and Mustar, P. (2001), *Research and innovation policies in the new global economy: An international comparative analysis*, Edward Elgar Publishing.

Li, Z., Wu, M. and Chen, B. R. (2017), 'Is road infrastructure investment in china excessive? evidence from productivity of firms', *Regional Science and Urban Economics* **65**, 116–126.

Lin, Y. (2017), 'Travel costs and urban specialization patterns: Evidence from china's high speed railway system', *Journal of Urban Economics* **98**, 98–123.

Martincus, C. V., Carballo, J. and Cusolito, A. (2017), 'Roads, exports and employment: Evidence from a developing country', *Journal of Development Economics* **125**, 21–39.

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

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

Moretti, E. and Wilson, D. J. (2014), 'State incentives for innovation, star scientists and jobs: Evidence from biotech', *Journal of Urban Economics* **79**, 20–38.

Nunn, N. and Puga, D. (2012), 'Ruggedness: The blessing of bad geography in africa', *Review of Economics and Statistics* **94**(1), 20–36.

Pakes, A. and Griliches, Z. (1980), 'Patents and r&d at the firm level: A first report', *Economics letters* **5**(4), 377–381.

Percoco, M. (2015), 'Highways, local economic structure and urban development', *Journal of Economic Geography* **16**(5), 1035–1054.

Redding, S. J. and Turner, M. A. (2015), Transportation costs and the spatial organization of economic activity, *in* 'Handbook of regional and urban economics', Vol. 5, Elsevier, pp. 1339–1398.

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

Romer, P. M. (1990), 'Endogenous technological change', *Journal of political Economy* **98**(5, Part 2), S71–S102.

Smith, K. (2005), 'Measuring innovation'.

Stock, J. H. and Yogo, M. (2005), 'Testing for weak instruments in linear iv regression, in dwk andrews and jh stock, eds., identification and inference for econometric models: Essays in honor of thomas j. rothenberg. cambridge: Cambridge university press'.

Talbert, R. J. (2000), *Barrington Atlas of the Greek and Roman World: Map-by-map Directory*, Vol. 1, Princeton University Press, Princeton.

Wang, X., Xie, Z., Zhang, X. and Huang, Y. (2018), 'Roads to innovation: Firm-level evidence from people's republic of china (prc)', *China Economic Review* **49**, 154–170.

Xu, H. and Nakajima, K. (2017), 'Highways and industrial development in the peripheral regions of china', *Papers in Regional Science* **96**(2), 325–356.

Yamasaki, J. (2017), 'Railroads, technology adoption, and modern economic development: Evidence from japan'.

Appendix A

In this section we strictly follow Agrawal et al. (2017) in explaining the theoretical innovation model underlying Equations 1 and 3.

The deterministic innovation level in a province, K_t^*, is related to the level of motorways, $Motorways_t$, trough the relation $K_t^* = A * Motorways_t^{\alpha}$.

Authors argue that the adjustment rate depends on how far a region is from the deterministic level of innovation. The innovation adjustment rate is defined as $K_{t+j} = {K_t^*}^{1-\gamma} K_t^{\gamma}$, with $0 < \gamma < 1$. Authors show that the level of innovation at time $t+j$ is equal to $K_{t+j} = BR_t^{\beta} K_t^{\gamma}$, where $\beta = \alpha(1-\gamma)$ and $B = A^{1-\gamma}$. Taking the log of this latter equation, Eq. (1) is obtained. The parameter of interest, β, describes the rate at which knowledge creation responds to motorways endowment.

Appendix B

Table B1: Main Results for the Impact of Motorways on the Innovative Activity, robustness to choice of the dependent variable, OLS and IV Estimations.

	(1)	(2)	(3)	(4)
	OLS	IV	OLS	IV
Dependent Variable: log Patents Fractional Count (per capita) 1988 based on Applicants Address				
$logMotorways_{p,1983}$	0.185***	0.136*	0.182***	0.163*
	(0.0499)	(0.0818)	(0.0529)	(0.0850)
$SpatialMotorways_{p,1983}$			0.115	-0.802*
			(0.516)	(0.466)
Geography	YES	YES	YES	YES
Applicants	YES	YES	YES	YES
Observations	89	89	89	89
R-squared	0.642	0.638	0.642	0.626
F-statistic		21.51		10.95

Notes: All specifications include the lagged dependent variable. Applicant controls include the log of the applicants (per capita) in each NUTS-3 region in 1983. Geography controls include surface, terrain ruggedness and elevation. We add one to all patent, applicants and motorways before to taking the log to include observations with value of zero. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Development and Governance in the Ex-Soviet Union: An Empirical Inquiry[1]

1 Introduction

The countries of the former Soviet Union over the last 20 years have experienced an impressive transformation. Massive and fundamental changes have occurred in almost all their political, economic and social aspects (Rechel et al., 2014). The transformation from command to market-oriented economies and the emergence of multi-parties political regimes in the former Soviet Union, against the background of the global processes of change, created an unstable political climate in North and Central Asia, with relevant elements of corruption. Slow economic recovery and social policy shortcomings launched numerous challenges for democracy and good governance in the rubble of the Soviet Empire (Petrovsky, 2004). The market economy and new freedoms positively impacted citizens' life but the levels of development within the ex-Soviet countries are still very different, as witnessed by life expectancy, which, among these countries, has a variation of seven years and half (Rechel et al., 2014). For these reasons, the economic and social performance of former Soviet Union is a fascinating field of study.

In this paper, we focus on two different dimensions of the former Soviet Union: development and governance. In order to analyse the development of this area, we decide to construct a composite index that is able to analyse the well-being from a more comprehensive point of view compared to the traditional GDP per-capita. Following the Bretton Woods conference in 1944, GDP became the main tool for measuring the economy of a country, gaining ever-greater importance in literature in the fifties and sixties. One of the main aims of macroeconomic policies over the last fifty years has been economic growth, measured as increases in Gross Domestic Product (Ivaldi, Bonatti and Soliani, 2016; Ivaldi, Soliani and Bonatti, 2016).

However, concerns over the long-term sustainability of economic growth have increased over the years and a shared vision that analyses GDP critically have been developed, especially when it is considered as the only point of reference to measure economic and social performance. Since the 1960s, the use and implicit interpretation of GDP (per capita) as a proxy of social welfare has received much criticism (van den Bergh, 2010). This criticism has come from some of the most respected economists of the 20th century, including various Nobel laureates, from Kuznets

[1]Joint work with Enrico Ivaldi and Riccardo Soliani

et al. (1941), Hicks (1939) and Samuelson (1961) to Arrow et al. (1996), Dasgupta (2000), Dasgupta and Mäler (2000) and Kahneman et al. (2004). As Bleys (2012) points out, both the potential pitfalls of macro-economic policies focused on stimulating economic growth and the problems involved in using GDP as a measure of well-being or economic welfare, has been recognized by economists and researchers from other social sciences. Crucial is the conclusion to which van den Bergh (2007) comes: "it is rational to dismiss GDP as an indicator to monitor economic progress and to guide public policy".

In order to overcome the mere income related dimension, a need of new tools for measuring a multidimensional well-being has raised, since there is not a single universal definition of well-being, nor a unique method to measure it. Its multidimensional nature makes it harder to assess, from which a number of theoretical, methodological and empirical issues emerge, the concept of development and well-being remains problematic and entailing a difficult selection concerning the variables to be included in its measurement. A turning point in the debate on the measurement of well-being is Sen's capability approach that identify welfare with "a person's capability to have various functioning vectors and to enjoy the corresponding well-being achievements" (Sen, 1985). This approach, supported by a wide range of literature (e.g. Sen, 1980, 1985, 1993, 1997, 2003, 2005)), is recognized as one of the best theoretical framework to represent the idea of multidimensional well-being (Pareglio et al., 2007). The richness of this approach makes its application difficult: the greatest limitation of the theoretical framework designed by Sen is the difficulty in addressing its quantitative application since the "operationalizing" process requires the measurement of dimensions related to the deepest and controversial aspects of the human being (Grasso, 2002).

The first major attempt to translate this capabilities approach into a tractable ranking of nations came in the 1990 UNDP Human Development Report. The objective was to better capture the complexity of human life by providing a quantitative approach and combining various socio-economic indicators into a measure of human development (UNDP, 1990). Indeed, indicators of wealth that reflect the quantity of resources available to a society do not provide information about the allocation of those resources. Thus, it is no wonder that countries with similar average incomes can differ substantially when it comes to quality of life of individuals (Soubbotina and Sheram, 2000). The new Human Development Index (HDI) was created to emphasize that people and their capabilities should be the ultimate criteria for assessing the development of a country, not economic growth alone. Once constructed a generic index of human development as UNDP did, it has been possible to follow this path in order to create other measures of well-being. In fact, the complexity of the concept of well-being can be clarified by aggregating different indicators. This approach find its basis in Dasgupta and Weale (1992): *"Measures of quality of life can take one of* *two forms: they can reflect the constituents of well-beieng, or alternatively, the can be measures of the access people have to determinants of well-being. Indices of health, welfare, freedom of choice, and more broadly, basic liberties are instances of the first: those indices which reflect the availability of food, clothing, shelter, potable water, legal aid, education facilities, health care, resources devoted to national security, and income in general, are examples of the latter".*

Then, the framework constructed by Partha Dasgupta, is an important tool that allow to create indices of well-being by aggregating indicators that reflect different aspects of this dimension (Grasso, 2004). Following this path, we construct an index of social and economic development that is supposed to capture the level of development of the ex-soviet countries. Starting point for our analysis is the "World Development Indicators" database by the World Bank (WB), and in particular, a subset of this database that represents the "Popular Development Indicators".
Once analysed levels of development, aware of the need to analyse economic welfare in a broader sense then the GDP, our attention focused on level of governance in these countries. In fact, under Soviet rule, corruption was widespread and highly institutionalised: most public officials and citizens regularly engaged in illicit activities and informal rules about illicit behaviours existed. Corruption under Soviet rule built an informal system parallel to the formal structure of communist rule. Then, given the systematic nature of corruption, it was very unlikely that corruption would simply disappeared with the rest of the Soviet Union: in fact, it did not (Stefes, 2006). Since informal rules and norms are not easy to change as formal rules and they are highly resistant to changes in the environment (March and Olsen, 1986; North, 1990), it is interesting to analyse level of governance –and then levels of corruption- in these countries.

As Smith (2007) points out, governance can be considered as an end in itself or as a mean to development and reduction of poverty. In particular, poor governance is among the most important causes of state failure and underdevelopment. Therefore, reforms and innovation in administration are an important pre-requisite for development (Ciborra and Navarra, 2005). As an evidence, cross-country regressions persistently demonstrate large and statistically significant correlations between institutional variables and growth (Shirley, 2005). In particular Rodrik et al. (2004) found that the quality of institutions "trumps" everything else in determining development. We analyse governance using the Worldwide Governance Indicators (WGI) created by the World Bank. The WGI is a long-standing research project aimed to develop cross-country indicators of governance. They consists of six composite indicators: "Voice and Accountability", "Political Stability and Absence of Violence/Terrorism", "Government Effectiveness", "Regulatory Quality", "Rule of Law", and "Control of Corruption" (Kaufmann et al., 2011).

In the former Soviet Union, the transition to market economies spurred legal reforms necessary for domestic and international markets. Formerly public assets were sold cheaply to insiders, often at low price. Unfair practices and asset grabbing emerged, long as the need of better laws and institutions. As is known, compliance with contracts and respect for property rights are necessary in order for markets to develop (Acemoglu and Robinson, 2012). However, nowadays, the concern about the power that the state wielders of violence must have is a relevant issue in Russia, since officials with sufficient coactive power to protect property rights are able also to take them away (Post-Soviet Affairs, 2014). Correspondingly, in the push for financial liberalization, the inflows to emerging markets became volatile, precipitating the 1997 Asian financial crisis. However, Asia had sound fundamentals: budget surpluses, high savings, low inflation, stable currencies, free trade, and a vast private sector. Thus, the institutional framework, namely financial deregulation, was considered as the real weak point and reforming was the most urgent challenge. The necessity of understanding how governance is related to development in post-Soviet countries emerged and offered a major field of research (Ramanujam et al., 2012).

The aim of the present paper is twofold. First: we propose an index of economic and social development in 2015 (ESDI). Secondly, we elaborate an index of governance (ESGI) in 2015. The correlation between the two indices and that between each of them and other indicators (GDP growth and Inequality-Adjusted human development index (IHDI)), allow to put forward some conclusions about growth and its economic and social sustainability in Russian Federation and the other countries studied.

2 Data

Economists, non-governmental organizations and policy maker have addressed the issue of measuring development for many decades. Following Santos and Santos (2014), there are three approaches to measuring development that now co-exist:

- One approach considers that development can be measured with some specification of a monetary indicator: in this approach, the GNP per capita is a good proxy for well-being.
- The second approach states that GNP per capita has too many weaknesses as an indicator of well-being and that it does not always correlate well with development goals. Then, following the approach a series of social indicators should be used to measure development.
- In the third approach, there is also a need for a summary measure that combines different social indicators into a single number. This approach has given rise to the construction of composite indices of development.

In order to analyse development and governance in the Former Soviet Union, we decide to follow the third approach and to construct two composite indicators: Ex-Soviet Development Index (ESDI) and Ex-Soviet Governance Index (ESGI) starting from two databases of the WB.
In particular, we use the World development Indicators database to construct ESDI. It presents the most current and accurate global development data available and includes more than thousand indicators. In order to reduce the number of indicators to deal with, we start from a subset of this database: "The Most Popular Development Indicators", which includes 15 economic and social indicators that measure the differences between the countries. They belong to the following dimensions:

- Economic Policy and Debt,
- Policy,
- Financial Sector,
- Infrastructure,
- Environment,
- Health,
- Education,
- Social Protection and Labour,
- Public Sector.

We list in Table 1 the initial list of indicators we considered for ESDI to which we applied principal component analysis. We decided not to include the two variables GNI per capita and Inflation (GDP deflator), which explain, respectively, the same phenomenon of GDP per capita and Inflation (consumer prices), as confirmed by the correlation matrix (see Appendix A for full table).

In the full database, the WB includes between the most popular development indicators the GDP growth that we have excluded since we will evaluate the similarity between the trend in our ESDI and in GDP growth.

Concerning the governance measurement, we use data from the Worldwide Governance Indicators database. the importance of the WGI database and its variables has been widely discussed (e.g. Kaufmann et al., 2011). A number of papers have been written presenting the database (e.g. Kaufmann, Kraay and Zoido, 1999; Kaufmann, Kraay and Zoido-Lobatón, 1999). The variables from WGI have been used for example by Neumayer (2002); Hart et al. (2005); Das and Andriamananjara (2006); Jung (2006). In particular, as pointed out by Thomas (2010), previous literature used them to assess the relationships between aspects of governance and growth (Kraay and Kaufmann, 2002; Dollar and Kraay, 2003; Kaufmann and Kraay, 2003; Naudé, 2004; Méon and Sekkat, 2005).

Table 2 shows the variables taken from the WGI in order to construct the ESGI. All these variables constitute estimations by the WB: the estimate gives the country's score on the aggregate indicator, in units of a standard normal distribution, i.e. ranging from approximately -2.5 to 2.5.

Table 1: Variables selected from World Development Indicators (WB) subset "The Most Popular Development Indicators"

Variable	Topic	Description
GDP per capita (current US$)	Economic Policy & Debt	GDP per capita is gross domestic product divided by midyear population. GDP is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources. Data are in current U.S. dollars.
Exports of goods and services (% of GDP)	Economic Policy & Debt	Exports of goods and services represent the value of all goods and other market services provided to the rest of the world. They include the value of merchandise, freight, insurance, transport, travel, royalties, license fees, and other services, such as communication, construction, financial, information, business, personal, and government services. They exclude compensation of employees and investment income (formerly called factor services) and transfer payments.
GINI index (WB estimate)	Poverty	Gini index measures the extent to which the distribution of income (or, in some cases, consumption expenditure) among individuals or households within an economy deviates from a perfectly equal distribution. A Lorenz curve plots the cumulative percentages of total income received against the cumulative number of recipients, starting with the poorest individual or household. The Gini index measures the area between the Lorenz curve and a hypothetical line of absolute equality, expressed as a percentage of the maximum area under the line. Thus, a Gini index of 0 represents perfect equality, while an index of 100 implies perfect inequality.
Inflation, consumer prices (annual %)	Financial Sector	Inflation as measured by the consumer price index reflects the annual percentage change in the cost to the average consumer of acquiring a basket of goods and services that may be fixed or changed at specified intervals, such as yearly. The Laspeyres formula is generally used.
Internet users (per 100 people)	Infrastructure	Internet users are individuals who have used the Internet (from any location) in the last 12 months. Internet can be used via a computer, mobile phone, personal digital assistant, games machine, digital TV etc.

Continued on next page

Table 1 continued from previous page

Variable	Topic	Description
Imports of goods and services (% of GDP)	Economic Policy & Debt	Imports of goods and services represent the value of all goods and other market services received from the rest of the world. They include the value of merchandise, freight, insurance, transport, travel, royalties, license fees, and other services, such as communication, construction, financial, information, business, personal, and government services. They exclude compensation of employees and investment income (formerly called factor services) and transfer payments.
Life expectancy at birth, total (years)	Health	Life expectancy at birth indicates the number of years a newborn infant would live if prevailing patterns of mortality at the time of its birth were to stay the same throughout its life.
Literacy rate, adult total (% of people ages 15 and above)	Education	Adult literacy rate is the percentage of people ages 15 and above who can both read and write with understanding a short simple statement about their everyday life.
Unemployment, total (% of total labor force) (modelled ILO estimate)	Social Protection & Labor	Unemployment refers to the share of the labor force that is without work but available for and seeking employment.
Poverty headcount ratio at national poverty lines (% of population)	Poverty	National poverty headcount ratio is the percentage of the population living below the national poverty lines. National estimates are based on population-weighted subgroup estimates from household surveys.
Agriculture, value added (% of GDP)	Economic Policy & Debt	Agriculture corresponds to ISIC divisions 1-5 and includes forestry, hunting, and fishing, as well as cultivation of crops and livestock production. Value added is the net output of a sector after adding up all outputs and subtracting intermediate inputs. It is calculated without making deductions for depreciation of fabricated assets or depletion and degradation of natural resources. The origin of value added is determined by the International Standard Industrial Classification (ISIC), revision 3. Note: For VAB countries, gross value added at factor cost is used as the denominator.
CO2 emissions (metric tons per capita)	Environment	Carbon dioxide emissions are those stemming from the burning of fossil fuels and the manufacture of cement. They include carbon dioxide produced during consumption of solid, liquid, and gas fuels and gas flaring.
Central government debt, total (% of GDP)	Public Sector	Debt is the entire stock of direct government fixed-term contractual obligations to others outstanding on a particular date. It includes domestic and foreign liabilities such as currency and money deposits, securities other than shares, and loans. It is the gross amount of government liabilities reduced by the amount of equity and financial derivatives held by the government. Because debt is a stock rather than a flow, it is measured as of a given date, usually the last day of the fiscal year.

Table 2: Variables from the Worldwide Governance Indicator database

Variable	Source	Description
Control of Corruption: Estimate	World Bank	Control of Corruption captures perceptions of the extent to which elites and private interests exercise public power for private gain, including both petty and grand forms of corruption, as well as "capture" of the state.
Government Effectiveness: Estimate	World Bank	Government Effectiveness captures perceptions of the quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies.
Political Stability and Absence of Violence/Terrorism: Estimate	World Bank	Political Stability and Absence of Violence/Terrorism measures perceptions of the likelihood of political instability and/or politically-motivated violence, including terrorism.
Regulatory Quality: Estimate	World Bank	Regulatory Quality captures perceptions of the ability of the government to formulate and implement sound policies and regulations that permit and promote private sector development.
Rule of Law: Estimate	World Bank	Rule of Law captures perceptions of the extent to which agents have confidence in and abide by the rules of society, and in particular the quality of contract enforcement, property rights, the police, and the courts, as well as the likelihood of crime and violence.
Voice and Accountability: Estimate	World Bank	Voice and Accountability captures perceptions of the extent to which a country's citizens are able to participate in selecting their government, as well as freedom of expression, freedom of association, and a free media.

3 Methods

Different techniques may be implemented to build a composite measure, which includes different variables into a global index. The methodology we used creates our indices through factorial analysis of the selected variables. In this case, the factorial scores, which represent the collocation of each observed variable in the representational space determined by the extracted factors, can be used as values of the index. To construct our composite indices we follow three steps (Nardo, 2005) :

1. Choose indicators between the initial list (Table 1 and Table 2).
2. Assign weights to each indicator.
3. Aggregate indicators to construct a composite index.

To select indicators, we apply a principal component analysis that is one of the methods of extraction of factors which factorial analysis makes use. We opted for this methodology in order to decide which variables have to be used (Ivaldi, Bonatti and Soliani, 2016; Ivaldi, Soliani and Bonatti, 2016). In component analysis, we simply transform the original variables into a new set of linear combinations (principal components). With principal component analysis, we partition the total variance by first finding the linear combination of the variables that accounts for the maximum amount of variance:

$$y_1 = a_{11}x_1 + a_{12}x_2 + ... + a_{1p}p,$$

where y_1 is the first principal component. Then, the procedure finds a second linear combination, uncorrelated with the first components, such that it accounts for the next largest amount of variance (after the variance attributable to the first component has been removed) in the system. The equation of the second component is:

$$y_2 = a_{21}x_1 + a_{22}x_2 + ... + a_{2p}p$$

The procedure goes on in this way. Thus, the use of principal components allows creating a set of uncorrelated variables (the components) by transforming a set of correlated variables. It means that the Pearson correlation between components is equal to 0 (Pituch and Stevens, 2015).

In order to make factors more interpretable, a process of rotation is applied. The rotation brings about the reduction of the weight of the factors that were comparatively "lighter" in the first step of the analysis, along with the increase of the weight of the factors that were comparatively "heavier" (note that here the absolute value is concerned) (Abdi, 2003; Krzanowski and Marriott, 1994). Indeed, in a non-rotation solution any variable is explained by two or more common factors, whereas in a rotation solution any variable is explained by a single common factor (Johnson and Wichern, 2002). A number of analytic rotation methods have been developed and although these rotation methods differ in a number of respects, perhaps the most fundamental distinction that can be made is between orthogonal and oblique rotations (Fabrigar et al., 1999). Orthogonal rotations produce factors that are uncorrelated (i.e., maintain a 90° angle between axes). Oblique methods allow the factors to correlate (i.e., allow the X and Y-axes to assume a different angle than 90°). Traditionally, researchers have been guided to orthogonal rotation because uncorrelated factors are more easily interpretable (Osborne, 2015).

Gorsuch (1983) lists four different orthogonal methods (equamax, orthomax, quartimax, and varimax) and 15 different oblique methods. As he points out, if the simple structure is clear, any of the more popular procedures can be expected to lead to the same interpretations. He then recommends rotating with varimax (orthogonal) or promax (oblique). The application of the varimax rotation was deemed appropriate as each factor has high correlations with a smaller number of variables and low correlations with the other variables and this generally makes interpretation of resulting factor easier (Kaiser, 1960).

Once extraction and rotation have been carried out, it is important to select which variables are to be used in the composite indicator. This has been done taking simultaneously into consideration three selection criteria:

- *Kaiser criterion*: following this approach it necessary to retain all factors extracted which have an eigenvalue greater than one because smaller values relate to factors which can explain less than what a single variable can explain (Kaiser, 1960).
- *Explained variance criterion*: in this case, the basis for the selection is the cumulative explained variance. A level of explained variance of 65% - 70% is considered significant (Stevens, 2012)
- *Scree test*: this method aims to give a graphical representation of the factors to be taken into consideration. The graph shows the value of the eigenvalue on the vertical axis and the number of components on the horizontal axis. The eigenvalues are plotted as points connected by a single line. According to the Cattell method, the choice of factors should be limited to the point where there is a levelling in the slope of the line (Cattell, 1966).

Once variables are chosen, we have opted for equal weighting. Indeed, even though it would be desirable to assign different weights to the various factors considered, there is no reliable basis for doing this (Mayer and Jencks, 1989). However, this does not mean no weighting, because equal weighting does imply an implicit judgment on the weights being equal (Nardo, 2005).

The further step is to build the two indices, ESDI and ESGI, by applying factor analysis performed on the selected variables. In this case, we can use as index values the factorial scores, which represent the position of each observation in the space of representation identified from the extracted factors (Hogan and Tchernis, 2004). While several different methods of estimating factor scores are available, as Pituch and Stevens (2015) point out, two are commonly used: regression method and sum/average scores. In the regression method, regression weights are obtained and factor scores are created by multiplying each weight by respective observed variable, which is in z-score form. The second method is simpler and allows to sum or average scores across the observed variables that load highly on a given factor as observed in the pattern matrix. Nevertheless, in this case we opted for the regression method, which allows estimating the score on the common factor as a linear combination of the original variables (Pituch and Stevens, 2015). Once obtained factorial scores we have been able to rank countries with respect to both indices and to observe if those countries with high scores in ESDI are the same countries that have high scores in ESGI.

To complete our analysis, we perform a cluster analysis to perform a bivariate classification of the countries according to both ESGI and ESDI since a cluster analysis can be applied to group the information on constituencies (e.g. countries) (Nardo, 2005). As Berkhin (2006) points out, clustering is a division of data into groups of similar objects; they are similar between themselves, but are dissimilar to the elements of other groups. Each group, called cluster, consists of objects that are similar between themselves and dissimilar to objects of other groups. Traditionally clustering methods range from those that are largely heuristic to more formal procedures based on statistical models. Then, clustering techniques are broadly divided in hierarchical and partitioning (Fraley and Raftery, 1998).

Hierarchical clustering techniques may be subdivided into agglomerative methods, which proceed by a series of successive fusions of the n individuals into groups, and divisive methods, which separate the n individuals successively into finer groupings (Everett, 2011). An agglomerative clustering starts with one-point (singleton) clusters and recursively merges two or more most appropriate clusters. A divisive clustering starts with one cluster of all data points and recursively splits the most appropriate clusters. The process continues until a stopping criterion (frequently, the requested number k of clusters) is achieved (Berkhin, 2006).

Relocation methods move observations iteratively from one group to another, starting from an initial partition. The number of groups has to be specified in advance and typically does not change during the course of the iteration. The method *k-means* (MacQueen et al., 1967; Hartigan and Wong, 1979) is the most common relocation method (Everett, 2011). For clustering via mixture models, relocation techniques are usually based on the EM algorithm (Dempster et al., 1977).

A recurrent problem has always been the determination of the number of clusters, because the way to determine the number of clusters is difficult (Davies and Bouldin, 1979). Davies and Bouldin (1979) presented a cluster analysis parameter that have to be minimized in order to indicate natural partitions of data sets. Fraley and Raftery (1998) proposed using the Bayesian information criterion (BIC) as the statistic criterion for the EM clustering method. Banfield and Raftery (1993) suggested using the approximate weight of evidence (AWE) for their model-based hierarchical clustering.

In the present work we decide to apply a two-steps cluster procedure to split countries into classes. TwoStep Cluster Analysis is a cluster analysis algorithm available in Predictive Analytics SoftWare (PASW). The TwoStep clustering procedure, as its name suggests, involves two distinct stages: in the first phase, original cases are grouped into preclusters. In the second step, the preclusters are clustered using the hierarchical clustering algorithm. Since being introduced in version 11.5 of the Statistical Packages for the Social Sciences (SPSS), TwoStep cluster analysis has been increasingly utilised in a variety of fields (Tkaczynski, 2017). The SPSS TwoStep

Cluster extends the model-based distance measure used by Banfield and Raftery (1993) to situations with both continuous and categorical variables. It utilizes a two-step clustering approach similar to BIRCH (Zhang et al., 1996) and it provides the capability to automatically find the optimal number of clusters.

4 Results and Discussion

Our analysis takes into account only 13 ex-Soviet Republics, since we eliminated from our database two countries, Uzbekistan and Turkmenistan, whose available data are not sufficient. Full databases with countries and variables are presented in Appendix B. First, we select variables to build the composite index of development (ESDI).

Starting from the variables listed in Table 1, we use principal component analysis to evaluate which variables should be left following the three criteria highlighted above, i.e. Kaiser's method, scree test and explained variance criterion. Table 3 shows Rotated Component Matrix, Figure 1 scree plot and Table 4 the values of explained variance.

According to Table 3, no pattern is inside the factor, then we may not consider the factors separately.
Since the second component explains just 62% of the variance (Table 4, we decide to take into account all three components and not to exclude any variable, as suggested also by the other two methods: all eigenvalues are greater then 1 (Kaiser criterion) and also the Scree Plot in Figure 1 suggests this choice.

Table 3: Factor score ESDI

Rotated component matrix(a)			
	Component		
	1	2	3
GDP per capita (current US$)	0.923	-0.081	0.305
Exports of goods and services (% of GDP)	0.839	-0.165	-.300
1/GINI*1000	-0.231	0.526	0.035
1/Inflation. consumer prices (annual %)	0.779	-0.422	-0.122
Internet users (per 100 people)	0.875	0.041	0.303
Imports of goods and services of (% of GDP)	0.174	-0.326	-0.884
Life expectancy at birth total (years)	0.718	-0.421	-0.032
Adult literacy rate population 15+ years, both sexes	0.384	-0.466	0.697
1/Unemployment*100	0.018	0.946	-0.033
1/Poverty headcount ratio at national poverty lines(% of population)*100	0.005	0.725	0.476
1/Agriculture value added of GDP*100	0.870	-0.063	0.304
CO2emissions (metric tons per capita)	0.567	0.440	0.524
1/Central government debt total of GDP*100	0.725	0.025	-0.112

Extraction method: principal component analysis
Rotation method: varimax with Kaiser normalization
Rotation converged in 5 iterations

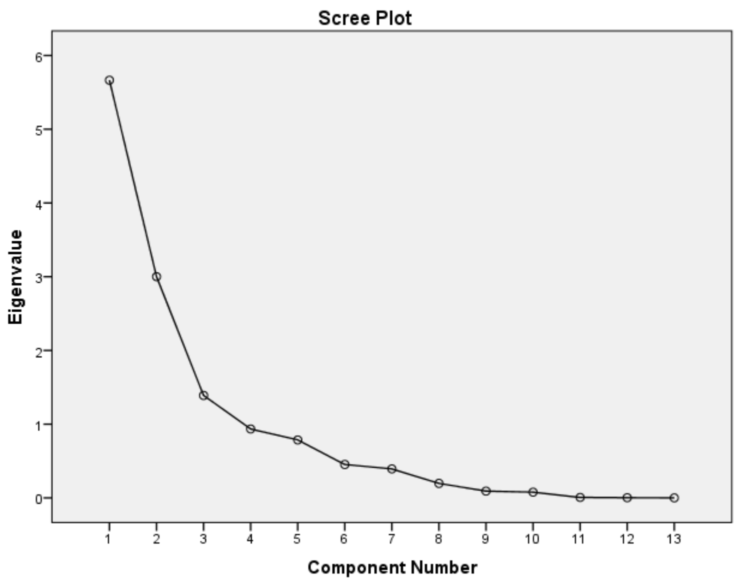
Then, we use principal component analysis to select variables in order to create the governance index (ESGI). As we can see from Figure 2, only the first component

Table 4: Total variance explained (development variables) ESDI

Components	Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %
1	5.281	40.624	40.624
2	2.608	20.064	60.688
3	2.166	16.664	77.352

Extraction method: principal component analysis

Figure 1: Scree plot for ESDI



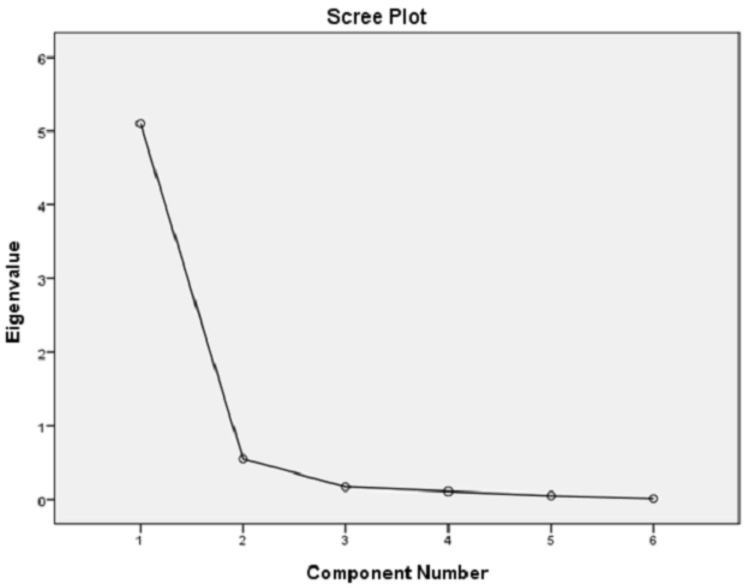
has an eigenvalue higher than 1: the first component fully explains the whole variability. It is no possible create rotated component matrix and total variance explained table; therefore, we include all the indicators in the model.

After deciding the variables of our model, we calculate the factorial scores, which represent the value of our composite indices. Note that factorial scores are compensative. Then, a country can show a satisfying performance even though it has a few bad indicators, provided they are counterbalanced by good records in the others.

Therefore, in some cases an analysis of one variable is needed, to grasp the real meaning of the result obtained. See for example the cases of Latvia in public debt ratio, or Moldova in unemployment rate.
Values of ESDI and ESGI are shown in Table 5 and 6, where we have ranked the countries from the better-performing to the less-performing ones.

The Ex-Soviet Republics Development Index takes into account both economic and social sustainability. Indeed, it includes macroeconomic variables (inflation, import, export, public debt), which express the economic constraints and sustainability of growth, and socio-economic variables (unemployment, poverty ratio, weight of agriculture interpreted as economic backwardness, internet users and the variables included in the IHDI: literacy, life expectancy, GDP per capita, Gini coefficient). Considering the quite low level of development of the ex-Soviet Republics, we decided to

Figure 2: Scree plot for ESGI



consider CO2 emissions not in terms of environmental sustainability, but rather as an indicator of industrial development (Ivaldi and Soliani, 2014).

Baltic countries are the best-performing. In the analysis of macroeconomic indicators, the three Republics show very similar characteristics (only with regard to the public debt, Estonia is better than the others, which however have a public $\frac{Debt}{GDP}$ ratio less than 50%), which set them nearer European Union (EU) countries, rather than the other Republics of the former Soviet Union. As to our indexes, Estonia, Latvia and Lithuania are at the top places in both classifications. This confirms the key role of fair political institutions to obtain economic and social development; the "European" path of development of Baltic Republics appears also from this point of view. Their strength is especially highlighted by the following variables: low inflation, low added value of agriculture, high GDP per capita, good education, export and many internet users. Estonia, the first of the Baltics, presents good performances in: public debt ratio (it is much better than in Lithuania and Latvia, which are, respectively, 10th and 11th), GDP, export, internet users and life expectancy, where it presents 3 years more than the other Baltics. Their weak point is the value of Gini index, since they are just before the Russian Federation and Georgia that are second-to-last and last. This weakness is confirmed by the variable "Poverty headcount ratio" and by the unemployment quite high. It seems that the market economy here is able to create good economic performance, with also some social satisfying results; but these countries, which express good governance, are not able yet to fight seriously and successfully against unemployment and inequality.

The performance of Russia, as to ESDI, is not entirely negative. It is at fourth place; but the distance from the first group, indicated by the ESDI score, is high (Latvia: 1.054; Russia: 0.059). We conclude that, apart from the Baltics, the eco-

Table 5: ESDI: scores and ranks

Rank	Country	Score
1	Estonia	2.226
2	Lithuania	1.390
3	Latvia	1.054
4	Russian Federation	0.059
5	Kazakhstan	-0.047
6	Georgia	-0.076
7	Belarus	-0.102
8	Armenia	-0.363
9	Azerbaijan	-0.426
10	Ukraine	-0.549
11	Moldova	-0.941
12	Kyrgyz Republic	-1.084
13	Tajikistan	-1.142

Table 6: ESGI: scores and ranks

Rank	Country	Score
1	Estonia	1.882
2	Lithuania	1.433
3	Latvia	1.307
4	Georgia	0.774
5	Armenia	-0.106
6	Moldova	-0.174
7	Kazakhstan	-0.356
8	Russian Federation	-0.630
9	Belarus	-0.661
10	Azerbaijan	-0.670
11	Kyrgyz Republic	-0.857
12	Ukraine	-0.858
13	Tajikistan	-1.082

nomic and social performance of the Russian economic system is in general better in comparison with the remaining countries of the former USSR, according to our compensative index. The Russian Federation follows the Baltics for number of internet users; this could be due to its big urban agglomerates with high density of population. On the other hand, Russia is bad-performing for what concerns inflation, Gini index and life expectancy: it emerges a large country with advanced aspects, like internet diffusion, and black spots, like life expectancy. The turbulence of the unequal Russian society pushes inflation. The mix of inflation, inequality and export-led growth, with the export based on oil and natural gas, reveals the fragility of the country. Such a weakness emerged in the crisis of December 2014, when the fall in the price of oil and Western financial sanctions, together with the widespread corruption, caused panic and the sharp contraction in the Russian economy in 2015 (Åslund, 2015). On the other hand, a recent inquiry, using the item–count technique, has obtained evidence that the very high Putin's approval ratings largely reflect the attitudes of Russian citizens. Much of the support for the leader is genuine and not the consequence of social desirability bias (Frye et al., 2017). The poor score obtained in ESGI (Russia rates eighth), apparently in sharp contrast with the support to Putin, further witnesses the major difficulties of establishing a market economy with a fair and effective democratic regime. On the contrary, it has been argued recently that Russia and others post-Soviet Republics have no incentive to improve institutions and prefer a low quality governance to maintain the monopoly of political rent as a precondition for extraction of economic rent (Melville and Mironyuk, 2016). At present, the Russian Federation is oriented to Europe as exporter but it does not import the European models of governance and development.

Kazakhstan, which in ESDI rank rates fifth, exports great amount of fuels and mineral products; its good performance is probably due to it. Commodities accounted for 88% of exports in 2014. But this lack of economic diversification makes the economy highly vulnerable to the global demand for commodities and their price. It is exposed to the business cycle of Russia, its main trading partner. Kazakhstan is followed by Georgia and Belarus. In the recent past, Georgia had several years of robust growth, but its macroeconomic outlook has deteriorated in recent years, because of the political fragility. Even if concerns about domestic instability have faded after the 2012 elections, Georgia is not returning to the previous model of growth based on the utilization of financial inflows to sustain domestic growth. Belarus' economic activity continues to stagnate, driven by the contraction in Russia, lower export revenues, and domestic structural problem. Its rank, then, confirms the fact that the Belarusian economy entered a recession in 2015, the first since 1995. Armenia occupies the eighth position of the rank, and its bad performance indicates that this country is at an inflection point: from 2000, growth slowed, and with it, the rate of poverty reduction. Some deceleration is due to the weaker performance of the country's main trading partners and a slowdown in remittance-providing countries. Azerbaijan rates ninth. The country is facing critical challenges due to the fall in oil prices, high inflation, and the crisis in the financial sector.

Then, there is Ukraine where disposable incomes contracted significantly in 2015 from the deep recession and high inflation. Moderate poverty increased from 15 percent in 2014 to 22 percent in 2015, while the poverty rate increased from 3.3 percent in 2014 to 5.8 percent in 2015 (World Bank).

The last three positions in ESDI rank are occupied by Moldova, Kyrgyz Republic and Tajikistan. They have almost the same position if we analyse GDP per capita. They are very bad-performing in number of internet users, Tajikistan in the last country for life expectation and Kyrgyz is the third from the bottom. In general, Moldova is quite better than the other two ex-Soviet Republics and in fact Moldova shows a higher factorial score that Tajikistan and Kyrgyz Republic. Even if Moldova is the tenth country for ESDI, it has made significant progress in reducing poverty and promoting inclusive growth since the early 2000s. The economy has expanded by an average of 5% yearly, driven by consumption and fuelled by remittances. The Kyrgyz Republic has progressively increased economic output over the past two decades, but the growth has been volatile. The economy remains characterized by significant informality and relies heavily on a few sectors and worker remittances from abroad. Finally, since independence, Tajikistan has done a remarkable job in reducing poverty, but progress has been slower in reducing non-monetary poverty (World Bank).

Analysing ESGI rank (Table 6) we find that the first three countries are once again the Baltics. In particular, Estonia is the first of the rank. If we observe the single variables ranks, it is easy to notice that Estonia is the first in each one except for political stability. Passing from ESDI to ESGI, we observe an improvement in ranking for Moldova, but transparency, accountability, and corruption are still crucial concerns. The former Soviet Union, as far as the thirteen Republics are concerned, and with regard to 2015, gives a quite large group of countries with poor administration and government, in comparison with the rest of Republics. Tajikistan occupies the last position and it is always the last in each variable. Ukraine is the second to last in ESGI rank; this is due to political instability, since the country has been interested by the civil war; this affects our results, which are based on 2015 data. We underline that all variables part of ESGI have similar trend and high correlation; therefore, we observe a homogeneity for each country, with reference to these variables.

Once analysed ESDI and ESGI separately, we test the well-supported hypothesis that good governance has a key positive relation with development and to do that we calculate correlation coefficient between ESDI and ESGI. The correlation between ESGI and ESDI is high: its value is almost 0.89. This confirms that governance and development are two sides of the same coin. Then looking at both of them we obtain a more complete and meaningful representation of the countries under scrutiny. If we remember that ESDI is an index of economic performance that includes also socio-economic variables, we can say that good governance lead to higher development in broad sense. Indeed, UN General Assembly in 2012 recognizes: "transparent, responsible, accountable, open and participatory government, responsive to the needs and aspirations of the people, is the foundation on which good governance rests, and that good governance at the national and international levels is essential for sustained economic growth, sustainable development and the eradication of poverty and hunger" (UN General Assembly, 2012). The blending of transparent, accountable and capable institutions of governance is essential for sustainable economic and social development.

Then, we introduce two new variables: GDP growth and HDI Adjusted (IHDI), and their corresponding ranks (Table 7). These variables allow us to test two further hypothesis:

1. Since we construct ESDI to enrich the mere income related dimension, we calculate the correlation between ESDI and GDP growth to observe if these two

measures provide different information.

2. Considering that ESGI comprises variables, not included in the IHDI, that represent human and civil development in a legal system, we want to test whether the former can be considered as an index that conveys similar information through different variables.

Table 7: Variables and corrisponding ranks

	ESDI	Rank	ESGI	Rank	GDP growth	Rank	IHDI	Rank
Armenia	-0.36	8	-0.11	5	3	3	0.73	10
Azerbaijan	-0.43	9	-0.67	10	1.1	8	0.75	7
Belarus	-0.10	7	-0.66	9	-3.89	12	0.8	4
Estonia	2.23	1	1.88	1	1.07	9	0.86	1
Georgia	-0.07	6	0.77	4	2.77	4	0.75	8
Kazakhstan	-0.04	5	-0.36	7	1.2	7	0.79	6
Kyrgyz Republic	-1.08	12	-0.86	11	3.47	2	0.66	12
Latvia	1.05	3	1.31	3	1.89	5	0.82	3
Lithuania	1.39	2	1.43	2	1.59	6	0.84	2
Moldova	-0.4	11	-0.17	6	-0.5	10	0.69	11
Russian Federation	0.06	4	-0.63	8	-3.73	11	0.8	5
Tajikistan	-1.14	13	-1.08	13	4.2	1	0.62	13
Ukraine	-0.55	10	-0.86	12	-9.9	13	0.75	9

The value of Rho calculated between GDP Growth and ESDI (0.049) demonstrates that there is no relation between these two variables. The concept of socio-economic performance and development expressed by ESDI cannot be reduced to the rate of growth. As Soubbotina and Sheram (2000), pointed out: although they reflect the average incomes in a country, GDP per capita have numerous limitations when it comes to measuring people's actual well-being. Thus, to judge the relative quality of life in different countries, one should also take into account other indicators: for instance, the distribution of income and incidence of poverty, people's health and longevity, access to education, and more. GDP growth is an imperfect proxy for welfare, or a means toward enhanced human development (Ranis, 2004). While growth in domestic product (GDP) is necessary to meet all essential human objectives, it is important to study how this growth translates - or fails to translate - into human development in various societies (UNDP, 1990). Correlation between ESGI and HDI Adjusted and Pearson's Rho assumes a value of almost 0.70. We remind that the IHDI combines country's average achievements in health, education and income; besides, it takes into account how those achievements are distributed among population by "discounting" each dimension's average value according to the level of inequality (UNDP, 2015). Then, ESGI can be considered as an index that conveys similar information with different variables. Finally, since we observed that correlation between ESDI and ESGI is high, we perform a bivariate two-steps cluster analysis to observe how countries can form different groups taking in account ESDI and ESGI simultaneously. Applying two-steps cluster analysis to ESDI and ESGI we obtain two classes where we can group the ex-soviet Republics (Table 8 and Figure 3).

Table 8: Cluster distribution

	N	% of combined	% of Total
Cluster			
1	3	23.1	23.1
2	10	76.9	76.9
Combined	13	100.0	100.0
Total			100.0

Figure 3: Cluster analysis results, MAP

Cluster analysis reveals a meaningful difference. The three Baltic Republics remain in the first cluster, separate from the rest of the former Soviet countries, placed on the second cluster. We can guess that this happens for both political and economic reasons. From 2004, Estonia, Lithuania and Latvia are members of the EU and also of NATO; differently from the other former Soviet republics, they do not belong to the Commonwealth of Independent States. All three states have adopted the Euro, Estonia since January 1, 2011, Latvia from January 1, 2014 and Lithuania since January 1, 2015. Moreover, Tallinn, Riga and Vilnius are increasingly oriented to the West, particularly Germany and Poland. They try to diversify their economies and look for alternative markets, to break the secular dependence from the Russian economy, today vulnerable to the crisis. In other words, they want to cut a close knot not only economic, but above all political. For example, the 2015 figures referring to the Latvia's import-export with Russia confirm this new trend. Indeed, they show 17% drop in export, compared to April 2014 and, above all, 23.7% fall in import, compared to May.

5 Concluding Remarks

The present paper proposes some considerations about the economic and social condition in the former USSR. We consider thirteen countries, bringing about a static analysis based on comparable and freely available data referring to 2015. The results offer suggestions for social and governance-related policies in ex-Soviet countries. In particular, our work can be useful to international observers to fully understand and identify possible misgovernance issues, corruption and institutional vulnerabilities among the countries considered, comparing governance and development performances. It may also allow European policy-makers and institutions to evaluate the potential of ex-Soviet countries, out of the European Union and currently under the Russian influence, to become more integrated partners and strengthen diplomatic relationships. Finally, our framework can be a tool to incentive governance to focus on poverty and development, concentrating on the multidimensionality of those two concepts.

The limitations of this analysis are those common to all sample studies. In this case, we compare very different countries. In further research, the results could be fruitfully analysed also at national level, or grouping the countries in macro-areas. The index does not include Uzbekistan and Turkmenistan, because we preferred to study full dataset, rather than dealing with missing data. Even though it strengthens the robustness of the analysis, this might be considered also a noticeable lack of complete information. Moreover, due to the intrinsically unobservable nature of the "true" level of governance in a country, any empirical measure of governance is just an imperfect proxy for the broader dimensions of governance that it reflects, as pointed out by Kaufmann et al. (2011). As a consequence, the estimates of governance, like the ESGI, are subject to non-trivial margins of error. Looking at a ranking, and making a qualitative analysis, rather than a strictly quantitative assessment, we tried to cope with this difficulty. However, our Index is supposed to be used to evaluate relative performances and not absolute values.

In the first part, we study a set of macroeconomic indicators that provide information about the sustainability of economic growth of the countries considered. Baltic Republics have a situation significantly better than the others do, and comparable to the EU. They are at the top places in both classifications, indeed. We can guess that fair political institutions allow them to obtain satisfying economic and social development, expressed by low inflation, low added value of agriculture, high GDP per capita, good education, export and many internet users. On the other hand, we must underline that inequality is a dramatic problem: their Gini index is at the bottom of ranking, just before Georgia and Russian Federation. Not surprisingly, also the variables related to poverty and unemployment are high. Both the link between inequality and unemployment, and the paradox of poverty in the wealth are confirmed. We can conclude that in Lithuania, Latvia and, particularly, Estonia, market economy offers a nice performance, with good social results in some cases; but these countries, nevertheless their good governance, are not able yet to reduce poverty, unemployment and inequality. These issue should be on the first page of their political agenda.

Looking at the ESDI index, we can observe that Ukraine suffers for the consequences of the civil war, which has engendered slumpflation and dramatic rise of poverty from 2014 to 2015. Also Armenia and Azerbaijan are marked by political and military tension. Armenia is at the eighth position. Since 2000, growth and poverty reduction have been slowing. This is due partly to the recession in its main trading partners. Azerbaijan rates ninth. The country is facing fall in oil prices, high inflation, and financial crisis.

Russian Federation and Kazakhstan base their export, and, largely, their economy, on fuels and mining products. Russia has a high number of internet users, but it is bad-performing for what concerns inflation, Gini index and life expectancy. The fragility of the country emerged in December 2014: the fall in the price of oil and Western financial sanctions from the Western countries for the Ukraine crisis, together with the widespread corruption, caused panic and the sharp contraction in the Russian economy in 2015 (Åslund, 2015). However, a recent inquiry has found evidence that the very high Putin's approval ratings reflect the real attitudes of Russian citizens (Frye et al., 2017). We can compare the result with the poor score obtained in ESGI. It witnesses the difficulties of establishing in Russia a market economy with a democratic regime. Our results depict a situation where the social tension, engendered by a model of growth led by the export of oil and gas that is jammed, pushes inflation and is contrasted by social policies, which hold unemployment quite low and create consensus to the autocracy in power. Kazakhstan, which in ESDI rank rates fifth, exports fuels and mineral products, and has a good performance; but the lack of economic diversification makes its economy highly vulnerable to the global demand for commodities.

Kazakhstan is followed by Georgia and Belarus. Georgia, after several years of robust growth, has seen its macroeconomic outlook deteriorated in recent years, due to its political fragility. Indeed, it is not returning to the previous model of growth based on the utilization of financial inflows to sustain domestic growth. Belarus' economic activity continues to stagnate. It suffers from the contraction in Russia, and has domestic structural problems. Export is decreasing. The Belarusian economy entered a recession in 2015, the first since 1995.

The last three positions in ESDI rank are occupied by Moldova, Kyrgyz Republic and Tajikistan, which have almost the same position of GDP per capita, but, in general, Moldova is quite better than the other two ex-Soviet Republics. Even if Moldova is the tenth country for ESDI, it has made significant progress in reducing poverty and promoting inclusive growth since the early 2000s, with an average growth rate of 5%. In Kyrgyz Republic the informal economy remains relevant and the country relies on few backward sectors and the worker remittances. Finally, since the independence Tajikistan has been reducing poverty, but is not able to develop social inclusion.

ESGI conveys information about the quality of governance. It has robust correlation with ESDI (around 71%), confirming the generally accepted relation between governance and socio-economic development. Remarkably, in this classification Russia falls in the last class, due to the weak points we have seen. On the other hand, Baltic Republics are at the top again; the result of the cluster analysis is further confirmed. Despite some investors' concerns regarding the geopolitical risks of the area, markets have seen further improvements in the economy and the arrival of new investors.

Interestingly, ESGI is correlated with IHDI: quality of life and good and fair institutions go together. On the contrary, ESDI is not correlated with GDP growth: growth and human development are not strictly linked each other.

The correlation between good governance and development, measured by ESGI and ESDI respectively, suggests that governance influences economic and social performance. However, correlation does not imply causal relation. We argue that in this case fair institutions allow development: the history of Baltic states in the last century, compared with Central Asia, suggests such an interpretation, in the wake of the well-known approach of Acemoglu and Robinson (2012); but in a long run analysis, for instance about the birth of property rights, we could single out the opposite causal relation. As implicit in Diamond's approach (Diamond and Renfrew, 1997), when an economic system is able to give surplus, the right of property rises and assumes the pivotal role in the legal structure of the state.

Before observing the single variables ranks, we must remember that all variables part of ESGI have similar trend and high correlation in each country. Now, we can easily notice that Estonia is the first in each one, except for political stability. In 2015, the thirteen Republics examined include a large group of countries with poor administration and government, in comparison with the rest of Republics. Passing from ESDI to ESGI, there is an improvement in ranking for Moldova; however, in that country transparency and accountability lack, and corruption is a serious problem. As to Tajikistan, not only does it occupy the last position, but it is also the last in each variable. Ukraine is the second to last in ESGI rank; this is due to its political instability, since the country has been interested by the civil war.

From our study, the portrait of a group of heterogeneous countries emerges. A small part of them looks at the EU and they records are really near it. Other countries have major internal weaknesses, due to political turmoils; not affordable, nor fair institutions; economic growth based almost exclusively on the export of fuels and raw materials, then weak and exposed to international cycle. Perhaps, our most impressive result regards Russia. It shows inequality, inflation and bad governance. According to our results, in spite of Putin's current popularity, Russian leading group could be jeopardised in the medium–long period.

References

Abdi, H. (2003), 'Partial least square regression (pls regression)', *Encyclopedia for research methods for the social sciences* **6**(4), 792–795.

Acemoglu, D. and Robinson, J. A. (2012), *Why nations fail: The origins of power, prosperity, and poverty*, Crown Books.

Arrow, K., Bolin, B., Costanza, R., Dasgupta, P., Folke, C., Holling, C., Jansson, B.-O., Levin, S., Mäler, K.-G., Perrings, C. et al. (1996), 'Economic growth, carrying capacity, and the environment', *Environment and Development Economics* **1**(1), 104–110.

Åslund, A. (2015), 'The current economic state of play in russia', *Gütersloh: Bertelsmann Foundation* .

Banfield, J. D. and Raftery, A. E. (1993), 'Model-based gaussian and non-gaussian clustering', *Biometrics* pp. 803–821.

Berkhin, P. (2006), A survey of clustering data mining techniques, *in* 'Grouping multidimensional data', Springer, pp. 25–71.

Bleys, B. (2012), 'Beyond gdp: Classifying alternative measures for progress', *Social Indicators Research* **109**(3), 355–376.

Cattell, R. B. (1966), 'The scree test for the number of factors', *Multivariate behavioral research* **1**(2), 245–276.

Ciborra, C. and Navarra, D. D. (2005), 'Good governance, development theory, and aid policy: Risks and challenges of e-government in jordan', *Information technology for development* **11**(2), 141–159.

Das, G. G. and Andriamananjara, S. (2006), 'Hub-and-spokes free trade agreements in the presence of technology spillovers: An application to the western hemisphere', *Review of World Economics* **142**(1), 33–66.

Dasgupta, P. (2000), 'Valutation and evaluation: Measuring the quality of life and evaluating policy', *Vol* .

Dasgupta, P. and Mäler, K.-G. (2000), 'Net national product, wealth, and social well-being', *Environment and development economics* **5**(1), 69–93.

Dasgupta, P. and Weale, M. (1992), 'On measuring the quality of life', *World development* **20**(1), 119–131.

Davies, D. L. and Bouldin, D. W. (1979), 'A cluster separation measure', *IEEE transactions on pattern analysis and machine intelligence* (2), 224–227.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977), 'Maximum likelihood from incomplete data via the em algorithm', *Journal of the Royal Statistical Society: Series B (Methodological)* **39**(1), 1–22.

Diamond, J. and Renfrew, C. (1997), 'Guns, germs, and steel: the fates of human societies', *Nature* **386**(6623), 339–339.

Dollar, D. and Kraay, A. (2003), 'Institutions, trade, and growth', *Journal of monetary economics* **50**(1), 133–162.

Everett, Brian, L. S. L. M. . S. D. (2011), *Cluster analysis*, Hoboken: Wiley.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C. and Strahan, E. J. (1999), 'Evaluating the use of exploratory factor analysis in psychological research.', *Psychological methods* **4**(3), 272.

Fraley, C. and Raftery, A. E. (1998), 'How many clusters? which clustering method? answers via model-based cluster analysis', *The computer journal* **41**(8), 578–588.

Frye, T., Gehlbach, S., Marquardt, K. L. and Reuter, O. J. (2017), 'Is putin's popularity real?', *Post-Soviet Affairs* **33**(1), 1–15.

Gorsuch, R. (1983), *Factor analysis (2nd ed.)*, Hillsdale: Lawrence Erlbaum Associates.

Grasso, M. (2002), 'Una misurazione del benessere nelle regioni italiane', *Politica economica* **18**(2), 261–292.

Grasso, M. (2004), The capability approach and sustainable development, *in* E. Tacchi, ed., 'Sustainability, Development and Environmental Risk', Foxwell & Davies.

Hart, D., Atkins, R. and Youniss, J. (2005), 'Knowledge, youth bulges, and rebellion', *Psychological Science* **16**(8), 661–662.

Hartigan, J. A. and Wong, M. A. (1979), 'Algorithm as 136: A k-means clustering algorithm', *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **28**(1), 100–108.

Hicks, J. R. (1939), 'Value and capital: An inquiry into some fundamental principles of economic theory'.

Hogan, J. W. and Tchernis, R. (2004), 'Bayesian factor analysis for spatially correlated data, with application to summarizing area-level material deprivation from census data', *Journal of the American Statistical Association* **99**(466), 314–324.

Ivaldi, E., Bonatti, G. and Soliani, R. (2016), 'The construction of a synthetic index comparing multidimensional well-being in the european union', *Social Indicators Research* **125**(2), 397–430.

Ivaldi, E. and Soliani, R. (2014), 'Is economical and social development really linked with the shift of labour from secondary to tertiary sector? evidence from the italian case', *International Journal of Economics and Finance* **6**(5), 26–32.

Ivaldi, E., Soliani, R. and Bonatti, G. (2016), *The measurement of well-being in the current era*, Nova Science Publishers Incorporated.

Johnson, R. A. and Wichern, D. W. (2002), *Applied multivariate statistical analysis*, Vol. 5, Prentice hall Upper Saddle River, NJ.

Jung, M. (2006), 'Host country attractiveness for cdm non-sink projects', *Energy Policy* **34**(15), 2173–2184.

Kahneman, D., Krueger, A. B., Schkade, D., Schwarz, N. and Stone, A. (2004), 'Toward national well-being accounts', *American Economic Review* **94**(2), 429–434.

Kaiser, H. F. (1960), 'The application of electronic computers to factor analysis', *Educational and psychological measurement* **20**(1), 141–151.

Kaufmann, D. and Kraay, A. (2003), 'Governance and growth: causality which way? evidence for the world, in brief', *World Bank, February* .

Kaufmann, D., Kraay, A. and Mastruzzi, M. (2011), 'The worldwide governance indicators: methodology and analytical issues', *Hague Journal on the Rule of Law* **3**(2), 220–246.

Kaufmann, D., Kraay, A. and Zoido-Lobatón, P. (1999), *Aggregating governance indicators*, Vol. 2195, world Bank publications.

Kaufmann, D., Kraay, A. and Zoido, P. (1999), 'Governance matters', *World Bank policy research working paper* (2196).

Kraay, A. and Kaufmann, D. (2002), *Growth without governance*, The World Bank.

Krzanowski, W. J. and Marriott, F. H. C. (1994), *Multivariate analysis*, Hoboken: Wiley.

Kuznets, S., Epstein, L. and Jenks, E. (1941), *National income and its composition, 1919-1938*, Vol. 1, National Bureau of Economic Research New York.

MacQueen, J. et al. (1967), Some methods for classification and analysis of multivariate observations, *in* 'Proceedings of the fifth Berkeley symposium on mathematical statistics and probability', Vol. 1, Oakland, CA, USA, pp. 281–297.

March, J. G. and Olsen, J. P. (1986), 'Popular sovereignty and the search for appropriate institutions', *Journal of Public Policy* **6**(4), 341–370.

Mayer, S. E. and Jencks, C. (1989), 'Poverty and the distribution of material hardship', *Journal of Human resources* pp. 88–114.

Melville, A. and Mironyuk, M. (2016), '"bad enough governance": state capacity and quality of institutions in post-soviet autocracies', *Post-Soviet Affairs* **32**(2), 132–151.

Méon, P.-G. and Sekkat, K. (2005), 'Does corruption grease or sand the wheels of growth?', *Public choice* **122**(1-2), 69–97.

Nardo, M., S. M. T. S. H. A. . G. E. (2005), *Growth without governance*, Paris: OECD Publishing.

Naudé, W. A. (2004), 'The effects of policy, institutions and geography on economic growth in africa: an econometric study based on cross-section and panel data', *Journal of International Development* **16**(6), 821–849.

Neumayer, E. (2002), 'Do democracies exhibit stronger international environmental commitment? a cross-country analysis', *Journal of peace research* **39**(2), 139–164.

North, D. C. (1990), *Institutions, institutional change and economic performance*, Cambridge university press.

Osborne, J. W. (2015), 'What is rotating in exploratory factor analysis?', *Practical Assessment, Research, and Evaluation* **20**(1), 2.

Pareglio, S., Grasso, M., Martinetti, E. C. et al. (2007), 'La qualità della vita nelle società avanzate: linee di ricerca per un modello basato sull'approccio delle capacità di amartya sen', *Economia e politica industriale* **2007**(1), 53–73.

Petrovsky, V. (2004), 'Growth, governance, and human development: Post-soviet transition in north and central asia', *Growth and governance in Asia* pp. 155–164.

Pituch, K. A. and Stevens, J. P. (2015), *Applied multivariate statistics for the social sciences: Analyses with SAS and IBM's SPSS*, Routledge.

Ramanujam, N., Verna, M. and Betts, J. (2012), 'Rule of law and economic development: A comparative analysis of approaches to economic development across the bric countries', *Available at SSRN: https://ssrn.com/abstract=2389074 or http://dx.doi.org/10.2139/ssrn.2389074* .

Ranis, G. (2004), 'Human development and economic growth', *Yale University Economic Growth Center Discussion Paper* (887).

Rechel, B., Richardson, E. and McKee, M. (2014), 'Trends in health systems in the former soviet countriesbernd rechel', *European Journal of Public Health* **24**(suppl_2).

Rodrik, D., Subramanian, A. and Trebbi, F. (2004), 'Institutions rule: the primacy of institutions over geography and integration in economic development', *Journal of economic growth* **9**(2), 131–165.

Samuelson, P. A. (1961), The evaluation of 'social income': Capital formation and wealth, *in* 'The theory of capital', Springer, pp. 32–57.

Santos, M. E. and Santos, G. (2014), 'Composite indices of development'.

Sen, A. (1980), 'Equality of what?', *The Tanner lecture on human values* **1**, 197–220.

Sen, A. (1985), 'Commodities and capabilities. amsterdam: North-holland'.

Sen, A. (1993), Capability and well-being, *in* M. Nussbaum and A. Sen, eds, 'The Quality of Life', Oxford, Clarenden Press.

Sen, A. (1997), 'From income inequality to economic inequality', *Southern Economic Journal* **64**(2), 384–401.

Sen, A. (2003), Development as capability expansion, *in* S. Fukuda-Parr and A. Shiva Kumar, eds, 'Readings in human development: Concepts, measures, and policies for a development paradigm', Oxford University Press.

Sen, A. (2005), 'Human rights and capabilities', *Journal of human development* **6**(2), 151–166.

Smith, B. C. (2007), *Good governance and development*, Macmillan International Higher Education.

Soubbotina, T. P. and Sheram, K. (2000), *Beyond economic growth: Meeting the challenges of global development*, World Bank Publications.

Stefes, C. H. (2006), *Understanding post-Soviet transitions: corruption, collusion and clientelism*, Springer.

Stevens, J. P. (2012), *Applied multivariate statistics for the social sciences*, Routledge.

Thomas, M. A. (2010), 'What do the worldwide governance indicators measure?', *The European Journal of Development Research* **22**(1), 31–54.

Tkaczynski, A. (2017), Segmentation using two-step cluster analysis, *in* 'Segmentation in social marketing', Springer, pp. 109–125.

UNDP (1990), *Human development report*, New York Oxford: Oxford University Press.

UNDP (2015), *Human development report 2015: Work for human development.*, New York: PBM Graphics.

van den Bergh, J. C. (2007), Abolishing gdp tinbergen institute discussion paper, Technical report, TI 2007-019/3, Vrije Universiteit Amsterdam, and Tinbergen Institute.

van den Bergh, J. C. (2010), 'Relax about gdp growth: implications for climate and crisis policies', *Journal of Cleaner Production* **6**(18), 540–543.

Zhang, T., Ramakrishnan, R. and Livny, M. (1996), 'Birch: an efficient data clustering method for very large databases', *ACM Sigmod Record* **25**(2), 103–114.

Appendix A: correlation matrix

Correlations

		GDP per capita (current US$)	Exports of goods and services (% of GDP)	GNI per capita, PPP (current international $)	GINI index (World Bank estimate)	Inflation, consumer prices (annual %)*100	Internet users (per 100 people)	Imports of goods and services (% of GDP)
GDP per capita (current US$)	Pearson Correlation	1	0.651(*)	0.948(**)	0.336	– 0.406	0.887(**)	– 0.066
	Sig. (2-tailed)		0.016	0.000	0.261	0.168	0.000	0.831
	N	13	13	13	13	13	13	13
Exports of goods and services (% of GDP)	Pearson Correlation	0.651(*)	1	0.527	0.129	– 0.082	0.672(*)	0.487
	Sig. (2-tailed)	0.016		0.064	0.674	0.790	0.012	0.092
	N	13	13	13	13	13	13	13
GNI per capita, PPP (current international $)	Pearson Correlation	0.948(**)	0.527	1	0.252	– 0.308	0.908(**)	– 0.317
	Sig. (2-tailed)	0.000	0.064		0.406	0.306	0.000	0.292
	N	13	13	13	13	13	13	13
GINI index (World Bank estimate)	Pearson Correlation	0.336	0.129	0.252	1	– 0.258	0.276	0.083
	Sig. (2-tailed)	0.261	0.674	0.406		0.395	0.361	0.786
	N	13	13	13	13	13	13	13
Inflation, consumer prices (annual %)*100	Pearson Correlation	– 0.406	– 0.082	– 0.308	– 0.258	1	– 0.251	– 0.184
	Sig. (2-tailed)	0.168	0.790	0.306	0.395		0.408	0.547
	N	13	13	13	13	13	13	13

Correlations

		GDP per capita (current US$)	Exports of goods and services (% of GDP)	GNI per capita, PPP (current international $)	GINI index (World Bank estimate)	Inflation, consumer prices (annual %)*100	Internet users (per 100 people)	Imports of goods and services (% of GDP)
Internet users (per 100 people)	Pearson Correlation	0.887(**)	0.672(*)	0.908(**)	0.276	− 0.251	1	− 0.206
	Sig. (2-tailed)	0.000	0.012	0.000	0.361	0.408		0.499
	N	13	13	13	13	13	13	13
Imports of goods and services (% of GDP)	Pearson Correlation	− 0.066	0.487	− 0.317	0.083	− 0.184	− 0.206	1
	Sig. (2-tailed)	0.831	0.092	0.292	0.786	0.547	0.499	
	N	13	13	13	13	13	13	13
Life expectancy at birth, total (years)	Pearson Correlation	0.618(*)	0.690(**)	0.454	0.373	− 0.389	0.613(*)	0.246
	Sig. (2-tailed)	0.025	0.009	0.119	0.209	0.189	0.026	0.419
	N	13	13	13	13	13	13	13
Adult literacy rate, population 15 + years, both sexes (%)	Pearson Correlation	0.575(*)	0.259	0.589(*)	0.196	− 0.103	0.482	− 0.348
	Sig. (2-tailed)	0.040	0.394	0.034	0.520	0.737	0.095	0.244
	N	13	13	13	13	13	13	13
Unemployment total (% of total labor force)	Pearson Correlation	− 0.077	0.005	− 0.216	0.399	− 0.248	− 0.176	0.214
	Sig. (2-tailed)	0.803	0.986	0.479	0.176	0.413	0.566	0.483
	N	13	13	13	13	13	13	13

Correlations

		GDP per capita (current US$)	Exports of goods and services (% of GDP)	GNI per capita, PPP (current international $)	GINI index (World Bank estimate)	Inflation, consumer prices (annual %)*100	Internet users (per 100 people)	Imports of goods and services (% of GDP)
Poverty headcount ratio at national poverty lines (% of population)	Pearson Correlation	− 0.097	− 0.107	− 0.322	0.355	− 0.374	− 0.331	0.492
	Sig. (2-tailed)	0.753	0.727	0.284	0.234	0.208	0.269	0.087
	N	13	13	13	13	13	13	13
Agriculture, value added (% of GDP)*100	Pearson Correlation	− 0.795(**)	− 0.596(*)	− 0.871(**)	− 0.183	0.186	− 0.903(**)	0.228
	Sig. (2-tailed)	0.001	0.032	0.000	0.550	0.544	0.000	0.454
	N	13	13	13	13	13	13	13
CO_2 emissions (metric tons per capita)	Pearson Correlation	0.637(*)	0.193	0.727(**)	0.107	0.091	0.581(*)	− 0.487
	Sig. (2-tailed)	0.019	0.527	0.005	0.729	0.767	0.037	0.092
	N	13	13	13	13	13	13	13
Central government debt, total (% of GDP)	Pearson Correlation	− 0.468	− 0.248	− 0.588(*)	0.033	− 0.032	− 0.703(**)	0.577(*)
	Sig. (2-tailed)	0.107	0.414	0.034	0.915	0.918	0.007	0.039
	N	13	13	13	13	13	13	13

Correlations

		Life expectancy at birth, total (years)	Adult literacy rate, population 15 + years, both sexes (%)	Unemployment total (% of total labor force)	Poverty headcount ratio at national poverty lines (% of population)	Agriculture, value added (% of GDP)*100	CO_2 emissions (metric tons per capita)	Central government debt, total (% of GDP)
GDP per capita (current US$)	Pearson Correlation	0.618(*)	0.575(*)	− 0.077	− 0.097	− 0.795(**)	0.637(*)	− 0.468
	Sig. (2-tailed)	0.025	0.040	0.803	0.753	0.001	0.019	0.107
	N	13	13	13	13	13	13	13
Exports of goods and services (% of GDP)	Pearson Correlation	0.690(**)	0.259	0.005	− 0.107	− 0.596(*)	0.193	− 0.248
	Sig. (2-tailed)	0.009	0.394	0.986	0.727	0.032	0.527	0.414
	N	13	13	13	13	13	13	13
GNI per capita, PPP (current international $)	Pearson Correlation	0.454	0.589(*)	− 0.216	− 0.322	− 0.871(**)	0.727(**)	− 0.588(*)
	Sig. (2-tailed)	0.119	0.034	0.479	0.284	0.000	0.005	0.034
	N	13	13	13	13	13	13	13
GINI index (World Bank estimate)	Pearson Correlation	0.373	0.196	0.399	0.355	− 0.183	0.107	0.033
	Sig. (2-tailed)	0.209	0.520	0.176	0.234	0.550	0.729	0.915
	N	13	13	13	13	13	13	13
Inflation, consumer prices (annual %)*100	Pearson Correlation	− 0.389	− 0.103	− 0.248	− 0.374	0.186	0.091	− 0.032
	Sig. (2-tailed)	0.189	0.737	0.413	0.208	0.544	0.767	0.918
	N	13	13	13	13	13	13	13
Internet users (per 100 people)	Pearson Correlation	0.613(*)	0.482	− 0.176	− 0.331	− 0.903(**)	0.581(*)	− 0.703(**)
	Sig. (2-tailed)	0.026	0.095	0.566	0.269	0.000	0.037	0.007
	N	13	13	13	13	13	13	13
Imports of goods and services (% of GDP)	Pearson Correlation	0.246	− 0.348	0.214	0.492	0.228	− 0.487	0.577(*)
	Sig. (2-tailed)	0.419	0.244	0.483	0.087	0.454	0.092	0.039
	N	13	13	13	13	13	13	13

Correlations

		Life expectancy at birth, total (years)	Adult literacy rate, population 15 + years, both sexes (%)	Unemployment total (% of total labor force)	Poverty headcount ratio at national poverty lines (% of population)	Agriculture, value added (% of GDP)*100	CO_2 emissions (metric tons per capita)	Central government debt, total (% of GDP)
Life expectancy at birth, total (years)	Pearson Correlation	1	0.435	0.447	0.136	− 0.426	0.191	− 0.393
	Sig. (2-tailed)		0.137	0.126	0.657	0.146	0.533	0.184
	N	13	13	13	13	13	13	13
Adult literacy rate, population 15 + years, both sexes (%)	Pearson Correlation	0.435	1	0.369	− 0.028	− 0.333	0.329	− 0.371
	Sig. (2-tailed)	0.137		0.215	0.928	0.265	0.273	0.212
	N	13	13	13	13	13	13	13
Unemployment total (% of total labor force)	Pearson Correlation	0.447	0.369	1	0.663(*)	0.359	− 0.476	0.310
	Sig. (2-tailed)	0.126	0.215		0.014	0.229	0.100	0.302
	N	13	13	13	13	13	13	13
Poverty headcount ratio at national poverty lines (% of population)	Pearson Correlation	0.136	− 0.028	0.663(*)	1	0.537	− 0.470	0.657(*)
	Sig. (2-tailed)	0.657	0.928	0.014		0.058	0.105	0.015
	N	13	13	13	13	13	13	13
Agriculture, value added (% of GDP)*100	Pearson Correlation	− 0.426	− 0.333	0.359	0.537	1	− 0.617(*)	0.609(*)
	Sig. (2-tailed)	0.146	0.265	0.229	0.058		0.025	0.027
	N	13	13	13	13	13	13	13
CO_2 emissions (metric tons per capita)	Pearson Correlation	0.191	0.329	− 0.476	− 0.470	− 0.617(*)	1	− 0.613(*)
	Sig. (2-tailed)	0.533	0.273	0.100	0.105	0.025		0.026
	N	13	13	13	13	13	13	13

Correlations

		Life expectancy at birth, total (years)	Adult literacy rate, population 15 + years, both sexes (%)	Unemployment total (% of total labor force)	Poverty headcount ratio at national poverty lines (% of population)	Agriculture, value added (% of GDP)*100	CO_2 emissions (metric tons per capita)	Central government debt, total (% of GDP)
Central government debt, total (% of GDP)	Pearson Correlation	− 0.393	− 0.371	0.310	0.657(*)	0.609(*)	− 0.613(*)	1
	Sig. (2-tailed)	0.184	0.212	0.302	0.015	0.027	0.026	
	N	13	13	13	13	13	13	13

*Correlation is significant at the 0.05 level (2-tailed)

**Correlation is significant at the 0.01 level (2-tailed)

Appendix B: Database

	GDP per capita (current US$)	Exports of goods and services (% of GDP)	GINI index (World Bank estimate)	Inflation, consumer prices (annual %)*100	Internet users (per 100 people)	Imports of goods and services (% of GDP)	Life expectancy at birth, total (years)
Armenia	3499.80	29.73	31.54	3.73	46.30	41.26	74.68
Azerbaijan	5496.34	37.81	16.64	4.17	61.00	34.82	70.76
Belarus	5740.46	60.07	26.01	13.53	59.02	59.12	72.98
Estonia	17,295.36	79.76	33.15	− 0.46	84.24	75.67	77.24
Georgia	3795.97	45.04	40.03	4.00	48.90	64.91	74.67
Kazakhstan	10,508.40	28.63	26.35	6.72	54.89	24.68	71.62
Kyrgyz Republic	1103.22	37.45	27.37	6.50	28.30	87.68	70.40
Latvia	13,664.94	58.76	35.48	0.20	75.83	60.18	74.19
Lithuania	14,172.22	77.29	35.15	− 0.88	72.13	77.41	73.97
Moldova	1843.24	43.43	28.53	9.68	46.60	73.73	71.46
Russian Federation	9057.11	29.53	41.59	15.53	70.52	21.21	70.37
Tajikistan	925.91	19.18	30.77	5.71	17.49	68.33	69.60
Ukraine	2114.95	52.77	24.55	48.72	43.40	54.76	71.19

	Adult literacy rate, population 15 + years, both sexes (%)	Unemployment total (% of total labor force)	Poverty headcount ratio at national poverty lines (% of population)	Agriculture, value added (% of GDP)*100	CO_2 emissions (metric tons per capita)	Central government debt, total (% of GDP)
Armenia	99.74	17.10	30.00	19.35	1.67	27.83
Azerbaijan	99.79	5.20	6.00	6.79	3.65	6.39

	Adult literacy rate, population 15 + years, both sexes (%)	Unemployment total (% of total labor force)	Poverty headcount ratio at national poverty lines (% of population)	Agriculture, value added (% of GDP)*100	CO_2 emissions (metric tons per capita)	Central government debt, total (% of GDP)
Belarus	99.62	5.90	5.10	7.80	6.68	25.22
Estonia	99.86	7.70	21.80	3.50	14.05	0.63
Georgia	99.75	13.40	14.80	9.19	2.05	32.53
Kazakhstan	99.73	4.10	2.80	5.01	15.81	13.22
Kyrgyz Republic	99.24	8.10	30.60	15.94	1.20	99.33
Latvia	99.90	10.00	22.50	3.27	3.79	41.59
Lithuania	99.82	11.30	19.10	3.44	4.54	43.76
Moldova	99.17	3.40	11.40	13.83	1.40	24.33
Russian Federation	99.68	5.10	13.40	4.63	12.65	12.68
Tajikistan	99.75	10.90	32.00	27.41	0.36	79.80
Ukraine	99.74	7.70	8.60	14.04	6.26	33.70

	Control of corruption: estimate	Government effectiveness: estimate	Political stability and absence of violence/terrorism: estimate	Regulatory quality: estimate	Rule of law: estimate	Voice and accountability: estimate
Armenia	− 0.44	− 0.17	− 0.21	0.22	− 0.32	− 0.55
Azerbaijan	− 0.92	− 0.34	− 0.50	− 0.29	− 0.61	− 1.44
Belarus	− 0.32	− 0.50	0.12	− 1.04	− 0.81	− 1.45
Estonia	1.27	1.05	0.76	1.67	1.36	1.17
Georgia	0.74	0.48	− 0.23	0.93	0.20	0.23
Kazakhstan	− 0.76	− 0.02	0.05	− 0.27	− 0.55	− 1.16
Kyrgyz Republic	− 1.11	− 0.84	− 0.78	− 0.42	− 0.94	− 0.53
Latvia	0.34	0.97	0.55	1.17	0.87	0.83
Lithuania	0.48	0.99	0.78	1.20	0.91	0.96
Moldova	− 0.85	− 0.38	− 0.10	0.02	− 0.27	− 0.02
Russian Federation	− 0.87	− 0.08	− 0.84	− 0.40	− 0.71	− 1.04
Tajikistan	− 1.00	− 0.75	− 0.68	− 1.01	− 0.96	− 1.44
Ukraine	− 1.00	− 0.38	− 1.93	− 0.63	− 0.79	− 0.08

Acknowledgement

Thanks to my supervisors, Anna and Maurizio, for their support over the years. I am grateful for the many things I have learned and the many experiences that this path has given me the opportunity to live.
Thanks to Enrico Rettore for every single piece of advice he gave me and for the time dedicated to me during my visiting at IRVAPP. Thanks also to all those who supported me during my stay in Trento, Sonia, Loris, Laura, Daniela and Olga.
Thanks to Enrico Ivaldi and Riccardo Soliani.
Thanks to my colleagues at PhD. Gianluca thanks for being a friend more than a colleague. Simone thanks for the profitable way we wrote our last work, always with a smile. Giulio thanks because LateX is now a wonderful tool for me. And thanks to Alessia, Carolina, Alessandro, Luigi and Chiara for all the days we spent together in the office.
Thanks to my family that always supports me in my choices and makes me believe in myself, in my abilities and always gives me the strength to dream.
Thanks to Fabrizio, you were always close to me, always reminding me to pull out the determination, even when I thought I didn't know how to do it anymore.
Thanks to the friends, all of them, because without friends the goals achieved would not be so good. Things shared always taste better. You have always been there, Giada, Chiara, Giulia, Federica, Stefania, Serena, Elisa, Veronica, Elena and Chiara, and you will always be there.
Thanks to all my dance mates. Valentina, as a good teacher, you have always taught me a lot, but what you always remind me is that you have to put yourself on the line and be tenacious. Chiara, Giulia and Vale, you are simply wonderful for the charge that only you can give. Dance brought us together and binds us forever. Thanks also to you little Sophi, who with your determination reminds us that it is right to believe in dreams.