How Can We Measure Sub-National Institutions? Some Ideas and Empirical Tests on European Regions, 1870-1910

Paul Caruana Galizia
The Legatum Institute, London

Abstract

Institutions are the dominant explanation of differences in economic performance between countries. One important criticism of institutional explanations is that institutions are national and thus cannot account for sub-national (regional) economic inequality. Another is that institutions cannot be empirically measured in a satisfactory manner. This paper offers some preliminary ideas in response to both criticisms. First, regional-level constraints on the executive are modelled as spatial-decay functions of national-level constraints on the executive. Second, the efficiency of institutions in distributing goods is modelled as a production frontier function. The ideas are empirically tested on industrialising Europe, when economic differentials between and within nations were widening. In putting forward two novel methods of institutional measurement and focusing on sub-national units of observation, this paper chiefly contributes to the debate on institutional measurement. In its empirical test, it makes use of a novel dataset on literacy rates for 199 European regions of seven countries over three benchmark years.

1. Introduction

Consensus in economics and economic history has shifted from explaining differences in economic performance as a function of differences in factor accumulation (Solow, 1956), to stressing the role of externalities from human and physical capital (Romer, 1986; Lucas, 1988), to emphasising differences in the allocation of resources to innovation (Romer, 1990). However, as North and

Thomas (1973: 2) put it early on, "the factors we have listed (innovation, economies of scale, education, capital accumulation, etc.) are not causes of growth; they are growth". Recent research has sought to move beyond the proximate causes of growth and tackle the fundamentals. In North and Thomas's (1973) view, it is differences in institutions that explain differences in growth. Acemoglu et al. (2005: 389), currently the leading proponents of the institutional view, argue that economic institutions are what matter for economic outcomes. These institutions collectively refer to the "… structure of property rights and the presence and perfection of markets". Put simply, economic institutions matter because they shape the structure of economic incentives in society. For this statement to be meaningful, the following conditions must be satisfied.

First, the term "institutions" is often all encompassing. As Voigt (2013) remarks, it has been used to refer to newspapers, supermarkets and even phone booths. We need to be clear about whether we are working with institutional structures (e.g. federal states) or organisations (e.g. the Church). Second, we need an empirical measure of institutions. To be convincing, theoretical arguments of institutions' importance must be supported empirically. Empirical measures to date often suffer in their imprecision. Third, we need to be clear on whether it is *de jure* or *de facto* institutions that matter or are being measured. For example, Italy's *national* institutions were, *de jure*, the same after unification, but we know that there were – and still are – *de facto* institutional differences within Italy, that is, subnational institutions (Dimico et al. 2012). This distinction matters. As Acemoglu and Robinson (2006: 325-6) write, *de jure* political power is "allocated by political institutions (such as constitutions or electoral systems)", while *de facto* political power "emerges from the ability to engage in collective action, or use brute force … lobbying or bribery". The two may work in tandem, but more often they are in conflict, as the case of Italy clearly shows (Tabellini, 2010).

This paper puts forward two proposals on how to measure institutions empirically. It is therefore not a paper on whether institutions matter for growth, but on how to measure institutions to begin with. Only after we do this can we develop arguments about the economic importance of institutions. The paper's main points are as follows:

1. Empirical measures of institutions should refer to more specific variables or concepts. The "rule of law" can be classified as an institution, but it is not representative of all institutions. This paper provides two different measures: sub-national executive constraints and the resource-distributive efficiency of institutions.
2. There should be a clear distinction between *de jure* and *de facto* institutions. The latter are usually harder to measure, but they matter more for economic performance (Acemoglu and Dell, 2010). This paper provides measures of *de facto* institutional variables, showing their within-country variation.
3. Many recent studies have used natural experiments or case studies that are helpful in the accumulation of evidence, but are not generalisable (Naritomi et al., 2012). This paper puts forward empirical estimation methods that can be applied across different countries and periods.

The data come from late-nineteenth-century Europe: a context in which countries were unifying, industrialising, and creating scope for domestic and cross-country differentials in institutional and economic outcomes. Again, the intention is to show that the methods put forward here can be applied across time and space: the sample consists of 199 regions from seven countries over 1870, 1900 and 1910. The goal is not to provide a definitive empirical measure or dataset, but to advance the debate in a concrete way. Hopefully, future work will examine, criticise and improve the measures proposed here. Indeed, the paper concludes with a list of practical and conceptual improvements that future work can pick up on.

2. Existing institutional measures

What are the existing measures of institutions? They centre on the ideas outlined in Acemoglu et al. (2005). Growth-promoting institutions 1) secure property rights and 2) allocate or distribute resources efficiently.

Rodrik et al. (2004), for example, use the World Bank's well-known 'Governance Matters' rule-of-law index (Kaufman et al., 2002). This composite index aggregates other indices that measure perceptions of the incidence of violent and non-violent crimes, the effectiveness and predictability of the judiciary, and contract enforceability, and ranges from -2.5 (weakest institutions) to 2.5 (strongest institutions). Basically, the index is a broadly conceived measure of the security of property rights. It is available for a sample of 79 countries but not at sub-national levels and cannot be used in historical work. The earliest annual data point is for 1996.

Overcoming this historical data limitation, Acemoglu et al. (2001) popularised the use of Gurr's (1997) Polity 'Constraints on the Executive' index. The indicator is recorded for a number of independent countries that enter and drop out of the time series, which starts in 1800. It rates countries on a scale of 7, with scores at the higher end indicating more constraints on the country's executive. A score of 1 indicates no constraints at all (for example, North Korea); 3 indicates slight to moderate constraints; 5, substantial constraints; and 7, executive parity or subordination. The remaining scores indicate intermediate constraints. This is a very useful measure of the security of property rights, but it is more commonly taken as a general index of a country's institutional environment. Acemoglu et al. (2003) show that it is correlated with other measures of institutions and with economic development. It has the advantage of being a long time series, but, again, it is not available at sub-national levels.

In a recent paper, Tabellini (2010: 678) provides executive constraint scores for a number of European regions for benchmark years between 1600 and 1850, by reviewing the historical literature on every sub-national region. His argument for doing so is "… the same institutions function very differently in different environments, suggesting that informal [de facto] institutions also play a role. The judicial system works very differently in southern and northern Italy …Yet the legal system … [has] been the same for 150 years …".

This newer work differs from that exemplified by Redding and Sturm (2008: 1782-3) who dismiss institutional explanations of disparities in economic performance among German cities inasmuch as "… [all] cities are part of the same country during all years of our sample, [so] there are no obvious differences in institutions between our treatment and control cities that could be responsible for the decline of the border cities". However, they never actually test for the possibility of sub-national institutions. In contrast, by going through the historical literature and assigning executive constraint scores to regions, Tabellini (2010) provides a very useful dataset that forms the basis of my first sub-national measure of institutions. The advantage of my technique over Tabellini's (2010) is that it can be easily extended to other countries over different years.

The reader will have noticed that while there is an abundance of data for different countries over various years on institutional variables that capture the property-rights side of things, there is a relative paucity with regard to the allocation and distribution of resources. My second measure seeks to fill this gap, but first I turn to the matter of extending Tabellini's (2010) work to my sample.

3. Proposal One: The Spatial Decay of Constraints on the Executive

In this section I propose estimating regional executive constraints as a spatial-decay function. That is, the farther away a region is from its country's seat of government – its national capital, the locus of the executive – the weaker will be its own executive constraints. "Executive constraints" is a popular institutional measure meant to capture the security of property rights from state expropriation (Acemoglu et al 2005).

This approach is rooted in recent political science literature but also harks back to much older ideas. Discussing the problem that slow communication posed for rulers in the early modern period, Braudel (1979: 326) famously remarked, "*L'espace, ennemi numéro one*"[1] Stasavage (2010) applies some empirical rigour to this concept, showing that the costs of distance prevented the regular assembly of governing officials in early modern Europe. More recently, Campante et al (2013) show that countries with isolated capitals are poorly governed. They provide some empirical support for the theory that executives are constrained by the threat of rebellion, and that this threat is less effective as distance from the seat of the executive increases. Michalopoulos and Papaioannou (2013: 1) explain within-country variation in modern-day African economic performance by the fact that "the explanatory power of national institutions rapidly decays for regions further from the capital cities", but they do not provide an explanation as to why this is so.

To further formalise the spatial-decay idea, I rely on Olsson and Hansson (2011), who similarly show that for a sample of ex-colonial countries the rule of law deteriorates the more peripheral a country's capital city is. Their paper outlines three points of departure for their argument, which I apply here.

1. Both executive and legislative power tend to originate almost exclusively from capitals. The rule of law is especially centralised when democracies are weak or nascent, as in late-nineteenth-century Europe's newly unified countries and weak democracies.
2. On the supply side, centralised executives must broadcast their power over space, to reach peripheral areas under their political control. Broadcasts are more effective over short distances. Herbst (2000) shows that in Africa public goods like the enforcement of property rights are most strongly felt around capitals among elite groups that control the government. Olsson and Hansson (2011) argue that institutions are public goods similar to knowledge or R&D, which has been empirically shown to be

[1] Usually translated as "Distance, [public] enemy number one".

highly geographically specific and to exhibit a spatial-decay function (Audretsch and Feldman, 1996).

3. On the demand side, even if we were to assume perfect broadcasting abilities of the executive, it is often the case that support for the executive and its laws decreases with distance from the capital. Similarly, Alesina and Spolaore (1997) assume that population size rather than geographical distance is the source of preference discord. Either way, distance from the capital – or seat of government – is negatively associated with the strength of law enforcement and with the willingness of local people to comply with the laws passed by the centralised executive.

3.1 An Empirical Analysis of the Spatial Decay of Executive Constraints

The above three points motivate an empirical implementation where executive constraints, located in a country's capital, decay over distance. So in my setting regions farther away from their capital city will exhibit lower executive constraints than those closed to it. This provides the following empirical model:

$$Inst_i = \alpha - \Theta D_{i-k} + \rho \mathbf{X}_i + \varepsilon_i \qquad (1)$$

where the dependent variable *Inst* is the executive constraints of region *i*, the regressor of interest *D* is the distance from region *i* to its country's capital *k*, from which the rule of law emanates. ***X*** is a matrix of control variables drawn from the relevant literature that are likely to affect institutional quality. The coefficient of interest Θ gives the rate at which a region's executive constraints respond to distance from the capital.

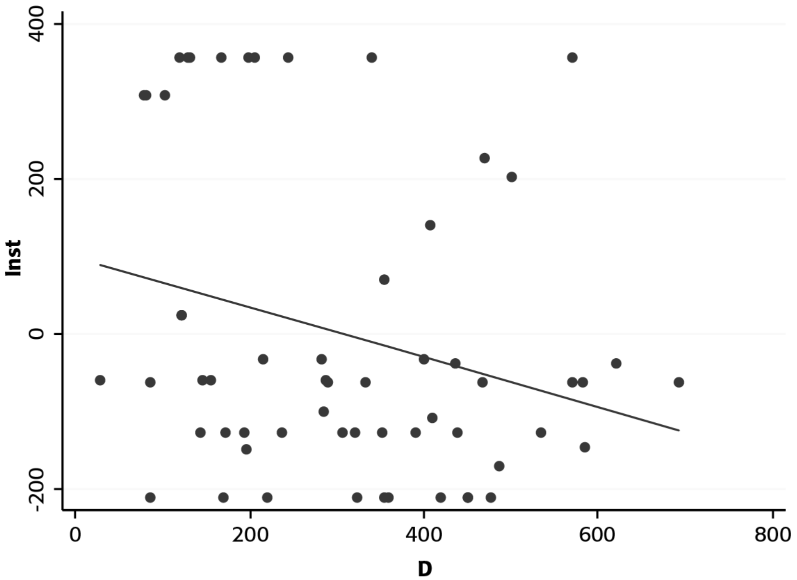
As noted above, Tabellini (2010) goes through the historical literature and assigns executive constraint scores for a sample of European regions for the benchmark years 1600, 1700, 1750, 1800, and 1850. He also extracts the first principal component of all these series, to arrive at a single historical "institutions" measure. This dataset framed my sample. I replicated Tabellini's map of European regions in a geographical information system (GIS), allowing me to measure *D* and other geographical controls. I took as my dependent variable – *Inst* – the principal component, to avoid the bias that could come from picking a particular year. Following Tabellini, I multiplied the component scores by 100, to enable them to be interpreted as percentages.

As for *D*, I measured the Euclidean distance from a region's centroid – its precise geographical centre – to the capital of its country. The last year in Tabellini's principal component is 1850, when Italy and Germany were not yet unified. This, however, is unlikely to affect my empirical analysis. First of all, Tabellini's coding of regions was done, as in Acemoglu et al. (2005), in 40-year windows around the benchmark years, so 1850 technically covers the period of national unification. Second, the nature of the principal component means that it captures a region's institutional heritage, which is something more meaningful, given what we know about the persistence of institutions, than a snapshot of its political situation. Lastly, most European countries were fiscally centralised by the mid-nineteenth century, and, as Dincecco (2009: 82) writes, "… fiscal centralisation undercuts provincial authority and contributes to the establishment of uniform rule".

The control variables I included in ***X*** are latitude, regional area, and a dummy that takes one for regions' whose *Inst* value drops below the sample median. The first is commonly used in the literature to capture the effects of physical geography (for example, climate or soil quality) on "institutional quality". The relationship has proved to be very robust across various samples and explains a high proportion of variation in institutional quality (Olsson and Hanson 2011). Regional size, while correlated with *D*, actually picks up distinct effects. It captures the overall ability of an executive to exercise its power across space. It is therefore an imperfect measure of what I am trying to do here, since, plausibly, a country's executive might be located too close to (far from) its regional centroids even though the region is large (small). It is the precise distance, *D*, that matters in this analysis. The dummy is particularly important as a control. If there is in fact a spatial-decay function to institutional quality, it might be better to be farther away from a corrupt or generally poorly constrained executive. The dummy variable controls for this effect on the estimation of Θ and, perhaps more interestingly, indicates whether it is indeed better to be far away from poorly constrained executives – in which case the coefficient on the dummy would be positive. It should be noted, however, that even when a central executive is poorly constrained or corrupt, distance from it might not necessarily be a good thing. Herbst (2000) and Michalopoulos and Papaioannou (2013) both show that distance from already corrupt and inefficient executive centres is associated with even more corruption. So it appears that a negative coefficient on the dummy variable is more likely than a positive one. I repeat, however, that the dummy's purpose is to get a more precise estimate of Θ.

Before delving into the empirical results, I show in Figure 1 the simple correlation between *Inst* and *D*. It is immediately clear that there is a strong negative relationship between the two variables.

FIGURE 1
Simple correlation between regional executive constraints and distance from capital city



The results of model (1) are shown in Table 1. The model fits the 58-region sample well, with an adjusted-R^2 of 65 per cent. That is, according to this statistic, two-thirds of the variation in regional institutional quality is explained by geographical variables alone (distance to capital, regional size, and latitude), along with a dummy variable that controls for below-median institutional quality. As an aside, this itself is interesting: one strand of the literature argues that geographical variables affect income only through their effects on institutions, while another argues for the direct effects of geography on income (Naritomi et al., 2012).

As expected from the conceptual discussion, the coefficient Θ on D, measuring Euclidean distance in kilometres from regional centroids to their capitals, is negative and significant at the 5 per cent

TABLE 1
Results of model (1)

	Inst
D	-0.255
R.S.E.	(0.095)**
Jack.	(0.094)**
Boot.	(0.091)**
Dummy	-205.99
R.S.E.	(35.322)***
Jack.	(34.397)***
Boot.	(34.433)***
Constant	-539.18
R.S.E.	(0.295)***
Jack.	(157.69)***
Boot.	(163.42)***
Controls	Latitude, Regional Size
Adj.-R^2	0.65
N	58

Inst is the first principal component of Tabellini's (2010) regional institutional measure, multiplied by 100. *D* is the Euclidean distance between regions' centroids and their country's capital city. *** denotes statistical significance at one per cent, and ** at five per cent. R.S.E. refers to robust standard errors clustered on regions. Boot. refers to bootstrapped standard errors, with 50 replications. Jack. refers to Jackknife standard errors, where the number of replications = N.

level. This tells us that there were significant costs to maintaining executive constraints over distances, and that peripheral regions were at a disadvantage in this regard. Olsson and Hansson (2011) get a similar result for their distance coefficient. Sticking to the national level, they measure the distance from a country's capital to its centroid, on which they regress a rule of law index. They find that a 1 per cent increase in their distance measure is associated with a 0.267 per cent decrease in their index, and a 0.123 per cent decrease when controlling for country size – values reassuringly close to the estimate of Θ achieved here, although their measure of "institutions" is different. As expected, the dummy variable – set at one for all regions whose *Inst* is below the sample median – is negative. This corroborates the idea that distance from the seat of the executive, even from a poorly constrained one, is more often than not a bad thing.

What does Θ really mean? Given that *Inst* is transformed into percentage values, Θ implies that for every kilometre increase in distance from region *i*'s centroid to its country's capital, its executive constraint score decreases by 0.00225 per cent. Using this coefficient, we can – to illustrate the technique's potential use – calculate regional values for Polity IV's executive constraints variable.

By way of example, Polity IV's dataset (the latest Polity dataset) assigns a score of 5 to France's (national) executive constraints in 1870. In accordance with the first point in my conceptual discussion, i.e. that both executive and legislative power originate almost exclusively from capitals and that the rule of law was especially centralised in nineteenth-century Europe, I assume this value to apply to the Paris (Seine) region. The centroid of Pyrénées-Orientales, a French department on the Mediterranean coast and bordering Spain, is some 795 kilometres from Paris. This implies that the *Inst* figure for Pyrénées-Orientales in 1870 is 3 (= 5 + (795 x -0.00225)), rounding to the nearest whole number. In 1900, when the Polity IV score for France is 7, the figure for Pyrénées-Orientales is 5 (= 7 + (795 x -0.00225)). I applied this method to my sample of 199 regions over three benchmark years (1870, 1900, and 1910). The results can be seen in the map in Figure 2, but the data only serve to show how the technique can be extended.

FIGURE 2
Map of regional executive constraints scores, 1-7.
Left to right: 1870, 1900, and 1910

Two obvious improvements come to mind for future research to consider when applying this technique to different samples. First, specification (1) needs to include more controls. Distance is lengthened by the presence of mountains or shortened by rivers. In an era when information was carried overland or by sea rather than transmitted through cables, this will matter even more, but it should not pose too great a problem for the method, given the wide availability of data on transport links (Caruana-Galizia and Marti-Henneberg, 2013) and on such other geographical variables as climate and terrain. The choice of controls will of course depend on the sample under study, so it is difficult to prescribe any here. Second, the coefficient Θ on distance *varies by country*. That is, each country is likely to have a different spatial-decay function. A "global" coefficient as in the calculations for Figure 2 obscures this, but the data used here are too sparse for a country-by-country estimation. Future research would do well to run estimations for different countries, allowing the coefficient to vary by country or whatever observational unit is under study. It would also be interesting to study *why* spatial-decay functions vary by country.

Two less substantive criticisms can be dealt with quickly here: measurement error and accuracy. Using centroids, the precise geographical centre of regions, to measure distance to capitals may be a source of measurement error. Possibly, a region's centroid does not necessarily reflect the point at which institutions matter. In Tabellini's (2010) regional groupings, for example, Sardinia and Sicily are placed together, meaning their centroid is a floating in the Tyrrhenian Sea. To check this potential bias, I used a GIS to plot 100 random points in each region, and used the average distance from the capital to each random point in place of the distance to centroid. The results of this experiment are reported in Table 2. The coefficient on the random point distance measure is -0.198 compared to -0.225 for the centroid measure. This 22 per cent gap may appear wide, but, in practice, it yields the same results when we calculated regional institutional values. Taking the Pyrénées-Orientales again but using this new coefficient, we get a value of 5 (= 7 + (795 x -0.00198)), the

Table 2
Results of model (1) with average distance to 100 random points

	Inst
D Random	-0.198
R.S.E.	(0.093)**
Jack.	(0.089)**
Boot.	(0.103)*
Dummy	-203.41
R.S.E.	(35.851)***
Jack.	(33.172)***
Boot.	(31.736)***
Constant	-541.18
R.S.E.	(157.01)***
Jack.	(155.55)***
Boot.	(183.51)**
Controls	Latitude, Regional Size
Adj.-R^2	0.64
N	58

Inst is the first principal component of Tabellini's (2010) regional institutional measure, multiplied by 100. *D* Random is the average Euclidean distance between 100 random points in a region, and its country's capital city. *** denotes statistical significance at one per cent, and ** at five per cent. R.S.E. refers to robust standard errors clustered on regions. Boot. refers to bootstrapped standard errors, with 50 replications. Jack. refers to Jackknife standard errors, where the number of replications = N.

same as when we used the centroid coefficient. The use of centroids or random points in calculating distance is immaterial.

To check how far off the results are from a given frame, I used it to replicate Tabellini's institutions measure for 1850. In this measure, Tabellini assigns Polity executive constraints scores to all the regions in his sample. I use the original country Polity IV scores, and apply the coefficient Θ from my model to them to arrive at regional institutional scores.[2] Using the original Polity IV dataset meant omitting Italy and Germany, two of Tabellini sample of eight countries, for which Polity IV does not provide scores (neither country was fully and officially unified by 1850). If this experiment results in regional scores that are close to Tabellini's, which were arrived at with a thorough reading of the historical literature, then it adds weight to the method's and model's reliability. The results of this exercise are shown in Table 3, which reports the unit difference between the replicated and actual executive constraint scores. The average difference between the two series is 1, the largest is 2. This may move a region slightly down the intermediate part of the scale, but no region undergoes a substantive shift, from, say, 1 (no constraints) to 7 (executive subordination).

While this check reassures us that the scores generated by the method fit with the literature and are well within bounds of "observed" scores, controlling for additional geographical features and estimating different country-coefficients remains an important step in validating this technique.

A more substantive question about this technique is whether the spatial-decay model is really suited to western European states. The spatial-decay-of-institutions explanation has more power for colonies (Michalopoulos and Papaioannou, 2013; Olsson and Hansson, 2011). In Europe, some cases do not fit the idea. For example, Galicia is farther from Madrid than is Catalonia, but it never posed

[2] The latest data can be found online at: http://www.systemicpeace.org/polity/polity4.htm.

TABLE 3
Replicating Tabellini's (2010) regional executive constraint scores for 1850

Region	Difference	Region	Difference
Galicia (ES)	1	Northern Ireland (GB)	0
Madrid (ES)	0	North (GB)	1
Asturias-Cantabria (ES)	1	North West (GB)	0
Pais Vasco (ES)	1	Wales (GB)	0
Navarra-Rioja (ES)	1	South West (GB)	0
Aragon (ES)	0	South East (GB)	0
Cataluna (ES)	1	East Anglia (GB)	0
Comunidad Valenciana (ES)	1	East Midlands (GB)	0
Balaeres (ES)	1	West Midlands (GB)	0
Murcia (ES)	1	Yorkshire Humberside (GB)	1
Andalucia (ES)	1	Mediterranean (FR)	1
Extremadura (ES)	1	South West (FR)	2
Castilla-La Mancha (ES)	0	South East (FR)	1
Castilla-Leon	0	Scotland (GB)	1
Norte (PT)	1	West (FR)	1
Centro (PT)	0	East (FR)	1
Lisboa (PT)	0	North (FR)	1
Algarve (PT)	0	Ile (FR)	0
Gronigen (NL)	0	Paris Basin (FR)	0
Oost (NL)	0	Regione Wallone (BE)	0
West (NL)	0	Vlaams Gewest (BE)	0
Zuid (NL)	0	Bruxelles (BE)	0

Difference is Tabellini (2010) executive constraint score for 1850 minus derived score for 1850. The average difference is 1; the largest is 2. German and Italian regions are not shown: Polity IV does not provide scores for these countries, because they were not fully and officially unified at the time.

a problem for Madrid whereas Catalonia did, and still does. We are always going to have some observations below and some above the regression line, as Figure 1 shows, but the general correlation is there, and while not perfect, it is strong enough to suggest some insights. The correlation was apparent to Braudel (1979), underpinning his idea that distance presented the *main* problem to governing elites, and has found recent empirical support in Stasavage (2010:

625), who, examining a broad sample of European representative assemblies, finds that in an era of costly communications and transport it was much easier to sustain an intensive form of political representation in geographically compact polities.

Sensu contrario: in this period of national unification, political elites were fully aware of the role played by transport networks in consolidating their rule across their countries. In Italy, for example, unaffordable railways were built, connecting obscure rural regions to one another. The railway industry was heavily subsidised by the new government, which also guaranteed high rates of return for investment in railway lines. Why? As Schram (2007: 3) writes: "For political reasons in the first decades after Italian unification, the role of the railways in the formation of the Italian state was considered to be very important. One minister of public works remarked in the 1860s that he was given the task of being the architect of Italian unity".

Put simply, if distance did not matter to *de jure* political power, then the Italian state would not have spent 13 per cent of its total budget, and 75 per cent of the public works budget, between 1861 and 1913 on railways (Schram 2007: 3). And this was not characteristic only of Italy. In France, Napoleon redrew the map of French *départements* to make them more accessible from Paris. The French state was also more directly involved in the design of the railway network (Schwartz et al., 2011; Mojica and Marti-Henneberg, 2011). The constitutional principle of *égalité* was applied in this policy, with national legislation dictating that each of the 95 *départements* should be connected by railway (Caron, 1997-2005). These projects tell us that distance mattered in the consolidation of *de jure* national political power.

Finally, none of what we have said clinches the case for rigid geographical determinism. Rather, these results show that distance was an important but not insurmountable or singular factor in the creation and maintenance of European political institutions.

4. Proposal Two: Estimating Regional Distributive Institutions

While property rights and rule-of-law-institutions have received considerable attention, much less work has been done on the role institutions play in allocating resources – on their distributive nature. As Acemoglu et al. (2005: 389-90) write, institutions "… not only determine the aggregate economic growth potential of the economy, but also an array of economic outcomes, including the distribution of resources in the future (i.e., the distribution of wealth, of physical or human capital). In other words, they influence not only the size of the aggregate pie, but how this pie is divided among different groups and individuals in society".

In a late-twentieth-century study, it is possible to proxy this "institution" using indicators of income inequality, education levels, and social mobility. Working in economic history, data unavailability makes this harder to do. It is harder still when we are seeking to unveil *de facto* regional distributive institutions rather than *de jure* national ones, for which we can look into tax policies, for example.

If we take the approach that these institutions determine the distribution of resources in an economy, and that efficient ones distribute resources more evenly, we can measure distributive institutional efficiency as the rate at which material resources are converted into resources like wealth and physical or human capital.[3] To borrow a term from Acemoglu et al (2005), we can measure how efficiently the "pie" is divided. One way this can be done empirically is by using a production frontier model.

4.1 An Empirical Analysis of Distributive Institutions

Technical efficiency in a production frontier model refers to the ability of a unit (regional distributive institution, in this case) to

[3] Moore et al. (1999) use a similar concept to measure governance in contemporary Africa, but implement it running a cross-section regression – human development index on GDP per capita – and then examining the residuals.

achieve maximum potential output (say, human capital) from given amounts of inputs (say, income). Farrel (1957) was perhaps the first to use frontier analysis to measure technical efficiency. Technically efficient units are those on the frontier, while inefficient ones, achieving less output than is technically possible, will be located below the frontier. Technical efficiency can thus be measured as the relationship between the realized output (Y) and what would be achieved if the unit were on the frontier (Y^*). That is, $0 \leq Y / Y^* \leq 1$.

The two main methods used in frontier analysis are data envelopment analysis (DEA) and stochastic frontier analysis. The former does not assume a functional form for the frontier (it is non-parametric), while the latter does; it is parametric, stochastic, and uses econometric methods. I use stochastic frontier analysis because it allows the frontier model's error term to be decomposed into statistical noise and inefficiency. DEA does not do this, and so ascribes deviations from the frontier only to inefficiency. In cross-sections of various regions like mine, this would make it hard to tell whether we are picking up heterogeneity or efficiency-differences between regions.

Stochastic frontier analysis was pioneered by Aigner et al. (1977). Having chosen a functional form for the production function, they put forward the following model: $y_i = f(x_i, \beta) + \varepsilon_i$, where y_i is the output achieved by region i, x_i is the vector of used inputs, β is a vector of parameters to be estimated, and ε_i is an error term composed of two elements, $\varepsilon_i = v_i + u_i$. The component v_i is the symmetric disturbance, capturing random variations in production due to random errors, observation errors and measurement errors, and is assumed to be identically and independently distributed as $N(0,\sigma_v^2)$. The component u_i is an asymmetric term, capturing technical inefficiency, and is assumed to be distributed independently of v_i, and to satisfy $u_i \leq 0$. Within this framework, maximum likelihood is used to produce consistent parameter estimates. The values for technical efficiency can be calculated as the expectation of the term u_i conditional on the composed error term ε_i. That is, technical efficiency for the i^{th} region is the ratio of observed output for the i^{th} region to potential output (frontier function). I use the standard assumption of a half-normal distribution for u_i. Therefore, technical efficiency (*TE*) can be measured as $TE_i = exp(-u_i)$, with $0 \leq TE_i \leq 1$ so as to ensure that scores remain either on or below the frontier. TE measures the efficiency of institutions in distributing resources (outputs) based on inputs. Very much in this vein, Nishimizu and Page (1982) argued that *TE* can be interpreted as a relative measure of managerial ability.

How do I adapt this model to my case? In other words, what do I use as an output and as an input to measure the efficiency of institutions? Acemoglu et al. (2005) mention, in particular, wealth and physical or human capital. For my current case, I do not have data on regional physical capital, but I do have, for the time being, data on regional literacy rates (a measure of human capital) and regional GDP per capita (a measure of income or economic prosperity).[4] Let me stress that I am not prescribing the use of these variables for future work. These data are what are presently available; the literature offers ample ideas as to what other variables can be used as inputs and outputs. Life expectancy, for example, can be used as an output, measuring average outcomes of healthcare provision, and government revenue can be used as an input, measuring the availability of public resources. The data are used to empirically illustrate the technique's potential.

Nevertheless, literacy rates are a useful indicator of how well distributed are gains from economic prosperity. Looking at Europe, we see that while countries took various paths to mass primary education, most had laws making it compulsory by the late nineteenth century: Italy in 1877, Britain in 1880, and France in 1882, for example (Soysal and Strang, 1989: 278). There are a number of explanations for this big education policy push during that period, but an important one is that elites came to see education as a way of improving a nation-state's international standing. For example, according to Sanderson (1999, p. 15-6), Baron Lyon Playfair, a Scottish

[4] Sources of literacy rates and GDP per capita are detailed in the appendix.

chemist and Liberal MP for the Universities of Edinburgh and St. Andrew's, later Postmaster General under Gladstone, opined that "if Britain were to remain without tariffs, as he [Playfair] advocated, then it would be self-defeating unless British industry were made as efficient as that of countries from which we imported. Education for economic efficiency was the only true protection".

This position was a response to the rise of industrial Germany, which had come to dominate the world chemicals and dye market by 1913. It was echoed by Sir Philip Magnus, an educational reformer, MP for the London University constituency, and director of the City and Guilds, who argued for "… state support for science and education to create industries from chemistry and electrical physics based on creatable and educable expertise …" (Sanderson, 1999: 15-6). GDP was the taxable base used to fund these plans and achieve these goals. As Engerman and Sokoloff (2000: 227-8) write referring to Latin America, "The institution of public primary schools … was the principal vehicle for high rates of literacy attainment and an important contributor to human capital formation…. Major investments in primary schooling did not generally occur in any Latin American country until the national governments provided the funds …".

Efficient regional distributive institutions are, then, institutions that channel economic prosperity (regional GDP per capita) into the production of more, and more widely distributed, resources (regional literacy rates). This chimes with the observation by Acemoglu et al. (2005: 390) that institutions "… influence not only the size of the aggregate pie, but how this pie is divided among different groups and individuals in society".

Now that I have selected my variables, I can proceed to specify the final implementation. Assuming a Cobb-Douglas production function, using linear presentation, and taking into account my variables, I re-write the function to be estimated as

$$\ln(LR_i) = \alpha + \beta \ln(GDPpc_i) + \varepsilon_i \qquad (2)$$

where i continues to index regions, LR is the literacy rate, and $GDPpc$

is GDP per capita. The parameter of interest β gives the elasticity of output (literacy) to input (GDP per capita). My sample covers 199 regions over three benchmark years, and to derive technical efficiency scores for each region at each year, I run the model as cross-sections for 1870, 1900, and 1910.

The results of model (2) are displayed in Table 4. Estimated values for the parameter β are highly significant across all years, indicating that the technical efficiency effect has an impact on output. That GDP per capita levels are positively associated with human capital is expected during this period. Attempts at improving literacy rates were mainly public initiatives and so needed large tax bases. That the effect gradually diminished over the period is also expected. Once large initial investments are made, it becomes easier to improve the literacy of posterity independently of public initiatives – for example, a parent who has been schooled can then teach his or her children to read and write before they enter school, if they ever do (Engerman and Sokoloff, 2000; Ljunberg and Nilsson, 2009).

The two variances of the two error components indicate that the inefficiency component u is much more statistically significant than

TABLE 4
Stochastic frontier model results

	1870	1900	1910
Constant	4.453	4.498	4.554
	(0.000)***	(0.000)***	(0.000)***
ln (GDP per capita)	0.015	0.014	0.006
	(0.000)***	(0.000)***	(0.000)***
$ln(\sigma_u^2)$	-0.700	-1.928	-2.458
	(0.000)***	(0.000)***	(0.000)***
$ln(\sigma_v^2)$	-36.413	-37.103	-37.626
	(270.941)	(270.812)	(270.360)
Log-likelihood	-75.132	47.723	100.676
N	200	200	200

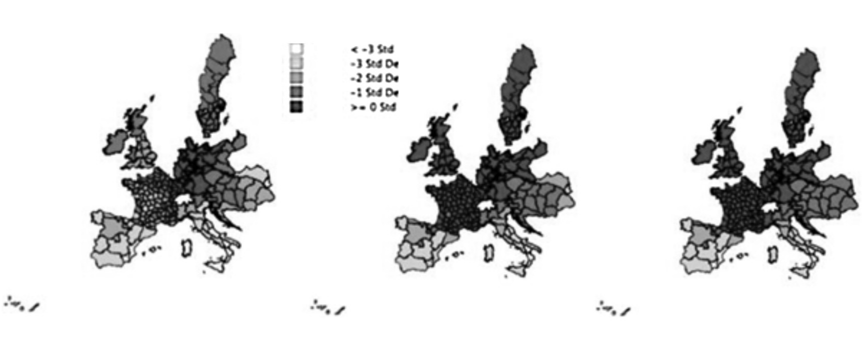
Estimated using maximum likelihood. Robust standard errors clustered on regions are in parentheses. *** denotes statistical significance at the one per cent level. Dependent variable is log of regional literacy rates.

the random component *v*. This implies that inefficiency *u* makes a larger contribution to the variability of the total error in the cross-sectional frontier model, and that inefficiency is highly significant across regions and years.

TE scores are shown in Figure 3. The mean technical efficiency score was 0.66 for 1870, 0.85 for 1900, and 0.89 for 1910. The standard deviations were 0.24 for 1870, 0.2 for 1900, and 0.17 for 1910. The main implication of these results is that regional institutions could have reduced their inputs (GDP per capita) by up to 34 per cent in 1870, 15 per cent in 1900, and 11 per cent in 1910, without reducing their output (literacy rates); simply by improving technical efficiency – or their *distributive institutional efficiency*. More efficient regional distributive institutions, therefore, could have higher levels of human capital without necessarily increasing their economic prosperity. Inefficient distribution quite clearly occurred, implying the presence of inefficient regional distributive institutions. While these data show the potential use of this approach, the following issues should be considered in future work.

Three possible improvements to this method immediately come to mind. First, the imperfect correlation between the two variables, whatever they may be, can be partially explained by other intermediating variables. Becker and Woessman (2011), for example, explain

FIGURE 3
Map of regional distributive institutional efficiency scores.
Left to right: 1870, 1900, and 1910



Prussian literacy rates in the nineteenth century as a function of Protestantism. Including such controls will result in a more precise estimate of error – or of the technical inefficiency – that is directly attributable to distributive institutions. Second, as mentioned earlier, the use of different input and output variables will not only act as a cross-check, but will also get us closer to a "true" value of distributive efficiency. One could also use aggregate indices of output or input – say, an output index that aggregates literacy and life expectancy. The possibility of using different variables as inputs and outputs is in itself an advantage of this method; it has broad applicability. Third, the nature of the relationship between the two variables must be considered. Here, literacy rates and per capita income are likely to be endogenous. This complicates interpretation of the results, since literacy rates may also be driving higher per capita incomes. Transforming the variables used, for example first-differencing, can alleviate endogeneity concerns, as can selecting variables with a strong underlying causal relationship. Still, this technique is not itself a causal model and so establishing exogeneity should not be the primary concern. Further, as Chang (2010) writes, there is in reality feedback between institutional set-ups and their outcomes. I have used a simple specification here to highlight the technique's potential use beyond this specific case.

Two other more fundamental issues may be more readily dealt with here. Why use a stochastic frontier model to begin with? Like the stochastic frontier approach, DEA assigns a score of 1 to a unit when comparisons with other units in the sample do not provide evidence of inefficiency in the use of input or output. A score of less than 1 indicates that a linear combination of other units can produce the same level of output with a lower level of input. Technically speaking, a DEA score reflects the radial distance from the estimated production frontier to the given unit. The most direct formulation of a DEA model, which I use here, is as follows. X_i is the vector of inputs (GDP per capita) into region *i*. Y_i is the corresponding vector of outputs (literacy rates). X_0 is the input into a region whose efficiency we want to determine, and Y_0 the outputs of that region. The

X's and the Y's are the data on GDP per capita and literacy rates, respectively. The following linear program gives the measure of efficiency for region 0:

$$\begin{aligned} & Min \quad \theta \\ & s.t. \quad \sum \lambda_i X_i \leq \theta X_0 \\ & \qquad \sum \lambda_i Y_i \geq Y_0 \\ & \qquad \lambda \geq 0 \end{aligned} \tag{3}$$

where λ_k is the weight given to unit i in its efforts to dominate unit 0 and Θ is the efficiency of unit 0. The λ's and Θ are the variables. Since unit 0 appears on the left of the equation as well, the optimal Θ cannot be more than 1. Solving this linear program yields the efficiency of region $0(\Theta)$, with $\Theta = 1$ meaning the region has efficient distributive institutions. Unlike the stochastic frontier approach, DEA does not distinguish between random error terms and inefficiency terms. This is why I chose the former as my primary method. This is partly why we cannot expect to see identical results in this comparative exercise. Still, there should be overall consistency between the two methods. Bauer et al. (1998) make some recommendations about how to compare the two approaches. In particular, efficiency scores obtained by the two approaches (*TE*, and Θ) should have comparable means and standard deviations. Furthermore, the two approaches should rank the regions according to their efficiency in approximately the same order. The results of the check are shown in Table 5. The means for each cross-section are almost equal, with the exception of the 1870 cross-section, where the difference between means is about 25 per cent. Given the differences of method, this is

TABLE 5
Mean and standard deviation of stochastic frontier and DEA models

	1870		1900		1910	
	TE	Θ	*TE*	Θ	*TE*	Θ
Mean	0.658	0.828	0.851	0.860	0.888	0.864
Stand. Dev.	0.238	0.101	0.195	0.052	0.168	0.55

TE refers to the technical efficiency score derived using the stochastic frontier model. Θ refers to the efficiency score derived using the DEA model.

a tolerable exception. Differences between standard deviations are larger but still consistent with those found in the literature (Iraizoz et al. 2003: 396). For the ranking check, I calculated the Spearman rank correlation between the two cross-sections of each year. For 1870, the correlation coefficient between the DEA and stochastic frontier cross-sections is 0.801, significant at the 1 per cent level. For 1900, it is 0.555, significant at the 1 per cent level; and for 1910, it is 0.536, significant at the 1 per cent level. The first correlation is by far the strongest, indicating very little difference in rankings. While the other two results are weaker, they are still strong correlations in an absolute sense and are also highly statistically significant. The conclusion is that the two methods produced similar results in terms of efficiency levels, as well as rankings of those levels.

Second, it is worth seeing whether the efficiency scores contain any useful information beyond their GDP per capita input. In their evaluation of the human development index (HDI) and its usefulness relative to per capita income, McGillivray and White (1993: 187) propose criteria for a variable's redundancy. First, a variable is redundant if the correlation coefficient is higher than 0.90 ("Level 1 Redundancy") or 0.70 ("Level 2 Redundancy"). Second, a variable is redundant if a "restricted" computation with the relevant component (in this case, GDP per capita) is excluded and remains highly correlated with excluded component. While I cannot test the second criterion, since the only component or input is GDP per capita, I can test the first. Table 6 shows that while the pair-wise and Spearman rank correlation coefficients are statistically significant, as we would

Table 6
Pair-wise and [Spearman] correlations of GDP per capita and technical efficiency scores

	1870	1900	1910
1870	0.340 [0.260]		
1900		0.640 [0.571]	
1910			0.572 [0.482]

Notes: All coefficients are significant at one per cent. Coefficients in brackets are Spearman rank coefficients; those not in brackets are pair-wise coefficients.

expect them to be, they pass the McGillivray and White (1993) "Level 1" and "Level 2" redundancy criteria. That is, the estimated technical efficiency scores contain useful information beyond their GDP per capita inputs.

5. Conclusions and Suggestions for Future Research

In this paper, I have used two different methods to construct two different regional institutional variables. First, I used a spatial-decay model to estimate the extent to which constraints on the executive deteriorate when moving outwards from capital cities, the seats of government. Second, I used stochastic frontier analysis to estimate the distributive quality of regional institutions. The production function took the level of economic prosperity in a region as inputs and human capital stocks as outputs. Efficient distributive institutions were those that produced high literacy rates with minimal levels of economic prosperity, indicating even and efficient distribution of resources.

In proposing these two methods, I sought to highlight two features of "institutions" that are too often ignored. First, there are important differences between *de jur*e and *de facto* institutions. Measuring *de jure* national-level institutions, using, say, a "rule of law" index, does not control for the presence of sub-national *de facto* institutions. Likewise, analysing a sample of sub-national units does not obviate the need to control for institutional variables. Second, there must be more precision about the nature of the institution under analysis and its effects. Indices like those on the rule of law are often taken as representative of an entire country's institutional environment. I have proposed two different measures of two different aspects of institutions. Regressing the executive constraints measure on the distributive efficiency measure, I get a coefficient of 1.5 and a constant of 3.5, both significant at 1 per cent, and an R^2 of 0.05. So while the two measures are positively correlated, the correlation is not tight. That is, they measure two different institutional qualities and so are likely to have different effects on economic performance.

While the proposed measures are not ideal, they do have advantages. They are both systematic, using a clear conceptual basis and historical data to obtain quantitative measures that can be used in an empirical test. They also have broad applicability. While I have used data on European regions, there is no reason why the methods cannot be applied anywhere else. The data requirements of the spatial-decay model are low, and those of the distributive model can be adapted to different contexts. This marks a break from papers that rely on case studies or natural experiments. This paper, however, did not aim to provide a dataset of regional "institutions" or to prescribe a definite empirical model. Rather, its objective was to show how certain empirical techniques can get us closer to measures and empirical tests of what is an important, nuanced explanation of economic performance.

Potential improvements on the techniques presented here are necessary and will move the debate forward. First, both models need to include more controls. The spatial-decay model needs to control for variables that lengthen distance, like mountains, or shorten it, like rivers. The frontier model can be easily adapted to control for other variables that may affect the imperfect correlation between the selected input and output measures. The choice of controls depends on the context under study. Second, the spatial-decay coefficient is, in reality, likely to vary by country and this should be allowed, given a greater dataset than the one used here. Differential decay rates are in themselves an interesting feature of the theory. Third, using different variables in the frontier model can act as a cross-check on the method. The choice of variables should also be sensitive to concerns about endogeneity. While it is not a causal model, an exogenous relationship will facilitate interpretation of its results.

Finally, it would be mistaken to interpret these results as a case for hard geographical determinism. The results show that distance was an important but not insurmountable variable shaping institutional quality. We have examples of regions that overcame the distance constraint, such as Galicia, politically stable though far from Madrid. These remind us that no single variable can explain all the variation in institutional quality. Future research would do well to explore precisely how certain regions overcame the distance constraint.

6. Notes

This research formed part of my doctoral research at the London School of Economics Department of Economic History, supervised by Max-Stephan Schulze and Nicholas Crafts; funded by the UK Economic and Social Research Council and an EU-Malta Government Steps Scholarship. I thank James Dowey, Jessica Magri, Max-Stephan Schulze, the anonymous reviewers of this paper, the editor, and seminar participants at a Caltech-All-UC-Huntington Library conference for discussions and comments. All errors are my own.

7. References

ACEMOGLU D., DELL M. (2010), "Productivity Differences Between and Within Countries", in *American Economic Journal*, Macroeconomics, 2(1), 169-188.

ACEMOGLU D., JOHNSON S., ROBINSON J.A. (2005), "The Rise of Europe: Atlantic Trade, Institutional Change and Economic Growth", in *American Economic Review*, 95, 546-579.

– (2001), "The Colonial Origins of Comparative Development: An Empirical Investigation", in *American Economic Review*, 91, 1369-1401.

ACEMOGLU D., JOHNSON S., ROBINSON J.A., THAICHAREON Y. (2003), "Institutional causes, macroeconomic symptoms: Volatility, crises, and growth", in *Journal of Monetary Economics*, 50, 49-123.

ACEMOGLU D., ROBINSON J.A. (2006), "De Facto Political Power and Institutional Persistence", in *American Economic Review*, 96(2): 325-330.

Aigner D.J., Knox Lovell C.A., Schmidt P. (1977), "Formulation and estimation of stochastic frontier production functions", in *Journal of Econometrics*, 6, 21-37.

Alesina A., Spolaore G. (1997), "On the number and size of nations", in *Quarterly Journal of Economics*, 112, 1027-1056.

Audretsch D.B., Feldman M.P. (1996), "R&D Spillovers and the geography of innovation and production", in *American Economic Review*, 86, 253-273.

Bauer P.W., Berger A.N., Ferrier G.D., Humphrey D.B. (1998), "Consistency conditions for regulatory analysis of financial institutions: a comparison of frontier efficiency methods", in *Journal of Economics and Business*, 50, 85-114.

Becker S., Woessmann L. (2009), "Was Weber Wrong? A Human Capital Theory of Protestant Economic History", in *Quarterly Journal of Economics*, 124(2), 531-596.

Braudel F. (1979), *La Méditerranée et le monde méditerranéen à l'époque de Philippe II*, Paris, Le Livre de Poche.

Campante F.R., Do Q-A., Guimaraes B.V. (2013), *Isolated Capital Cities and Misgovernance: Theory and Evidence*, National Bureau of Economic Research Working Paper 19028.

Caron F. (1997-2005), *Histoire des chemins de fer en France*, Paris, Fayard, 2 vols.

Caruana-Galizia P., Marti-Henneberg J. (2013), "European regional income and railways, 1870-1910: a preliminary report", in *Scandinavian Economic History Review*, 61(2), 167-196.

Chang H-J. (2011), "Institutions and economic development: theory, policy and history", in *Journal of Institutional Economics*, 7(2), 473-498.

Dimico A., Isopi A., Olsson O. (2012), *Origins of the Sicilian Mafia: The Market for Lemons*, University of Nottingham CREDIT Discussion Papers 12/01.

Dincecco M. (2009), "Fiscal Centralization, Limited Government, and Public Revenues in Europe, 1650-1913", in *Journal of Economic History*, 69, 48-103.

Engerman S., Sokoloff K. (2000), "History Lessons: Institutions, Factor Endowments, and Paths of Development in the New World", in *Journal of Economic Perspectives*, 14, 217-232.

FARRELL M.J. (1957), "The Measurement of Productive Efficiency", in *Journal of the Royal Statistical Society*, 120, 253-281.

GURR T.R. (1977), *Polity II: Political Structures and Regime Change, 1800-1986*, Unpublished Manuscript, University of Colorado, Boulder.

HERBST J. (2000), *States and Power in Africa: Comparative Lessons in Authority and Control*, Princeton, Princeton University Press.

IRAIZOZ B., RAPUN M., ZABALETA I. (2003), "Assessing the technical efficiency of horticultural production in Navarra, Spain", in *Agricultural Systems*, 78, 387-403.

KAUFMANN D., KRAAY A., ZOIDO-LOBATON P. (2002), *Governance Matters II - Updated Indicators for 2000/01*, World Bank Policy Research Department Working Paper No. 2772, Washington, D.C.

LJUNBERG J., NILSSON A. (2009), "Human capital and economic growth: Sweden 1870-2000", in *Cliometrica*, 3, 71-95.

LUCAS R.E. (1988), "On the mechanics of economic development", in *Journal of Monetary Economics*, 22, 3-42.

MCGILLIVRAY M., WHITE H. (1993), "Measuring Development? The UNDP's Human Development Index", in *Journal of International Development*, 5(2), 183-192.

MICHALOPOULOS S., PAPAIONNOU E. (2012), *Pre-colonial Ethnic Institutions and Contemporary African Development*, National Bureau of Economic Research Working Paper 18224.

MOJICA L., MARTI-HENNEBERG J. (2011), "Railways and population distribution: France, Spain and Portugal, 1870-2000", in *Journal of Interdisciplinary History*, 42 (1), 15-28.

MOORE M., LEAVY J., HOUTZAGER P., WHITE H. (1999), *Polity Qualities: How Governance Affects Poverty*, The Institute of Development Studies, University of Sussex.

NARITOMI J., SOARES R.R., ASSUNCAO J.J. (2012), "Institutional Development and Colonial Heritage within Brazil", in *Journal of Economic History*, 72(2), 393-422.

NISHIMIZU M., PAGE J.M. (1982), "Total Factor Productivity Growth, Technological Progress and Technical Efficiency Change: Dimensions of Productivity Change in Yugoslavia, 1965-78", in *The Economic Journal*, 92, 920-936.

Olsson O., Hansson G. (2011), "Country size and the rule of law: Resuscitating Montesquieu", in *European Economic Review*, 55, 613-629.

Redding S.J., Sturm D.M. (2008), "The Costs of Remoteness: Evidence from German Division and Reunification", in *American Economic Review*, 98, 1766-11797.

Rodrik D., Subramanian A., Trebbi F. (2004), "Institutions Rule: the Primacy of Institutions over Geography and Integration in Economic Development", in *Journal of Economic Growth*, 9, 131-165.

Romer P.M. (1986), "Increasing returns and long-run growth", in *Journal of Political Economy*, 94, 1002-1037.

– (1990), "Endogenous technical change", in *Journal of Political Economy*, 98, 71-102.

Sanderson M. (1999), *Education and Economic Decline in Britain, 1870 to the 1990s*, Cambridge University Press, Cambridge.

Schram A. (2007), *Railways and the Formation of the Italian State in the Nineteenth Century*, Cambridge, Cambridge University Press.

Schwartz R.M., Gregory I.N., Thevenin T. (2011), "Spatial history: railways, uneven development, and population change in France and Great Britain, 1850 1914", in *Journal of Interdisciplinary History*, 42 (1), 53–88.

Solow R.M. (1956), "A Contribution to the Theory of Economic Growth", in *Quarterly Journal of Economics*, 70(1), 65-94.

Soysal Y., Strang D. (1989), "Construction of the First Mass Education Systems in Nineteenth Century Europe", in *Sociology of Education*, 62(4): 277-288.

Stasavage D. (2010), "When Distance Mattered: Geographic Scale and the Development of European Representative Assemblies", in *American Political Science Review*, 104, 625-643.

Stephens W.B. (1998), *Education in Britain: 1750-1914*, London, St Martin's Press.

Tabellini G. (2010), "Culture and Institutions: Economic Development in the Regions of Europe", in *Journal of the European Economic Association*, 8, 677-716.

Voigt S. (2013), "How (Not) to Measure Institutions", in *Journal of Institutional Economics*, 9(1), 1-26.

8. Appendix

8.1 Literacy Data

There is heterogeneity in the sample and types of sources, but the aim was not to derive a precise figure for literacy but, rather, to approximate regional human capital levels, as an output for distributive institutional efficiency, in a way that would not bias the empirical results. There is no reason to think that a persistent bias – for example, literacy measured in one country being correlated with the empirical model's error term – exists in this dataset.

- *Austria-Hungary*. Regional indices (ratios relative to national rate) are from page 156 of Good D.F. (1984), *The Economic Rise of the Habsburg Empire, 1750-1914*, Berkeley, UC Press. I converted these into real rates, using the national rates on pages 127 and 118 in Cipolla C. (1969), *Literacy and Western Development*, London, Penguin.
- *Britain*. Rates for England and Wales are from volumes 33, 63, and 73 of the *Annual Report of the Registrar-General of Births, Deaths, and Marriages in England and Wales*, London, H.M. Stationery Office. These figures refer to the percentage of people unable to sign the marriage register. Rates for Ireland are from *Ireland Census 1911, General Report and Ireland Census 1901, General Report, Part 2*, London, H.M. Stationery Office. Scottish rates come from page 127 in Cipolla C. (1969), *Literacy and Western Development*, London, Penguin. The Scottish figure for 1910 is an average for Wales, England, and Ireland.
- *France*. Rates are from the census books. For 1870, *Statistique de la France 1872 – Tome XXI – Population*, Paris, Imprimerie Nationale. For 1901, *France Récensement 1901 – Tome IV – Résults Généraux*, Paris, Imprimerie Nationale. For 1901, *France Récense-*

ment 1901 – Tome IV – Résults Généraux, Paris, Imprimerie Nationale. For 1911, *France Récensement 1911 – Tome I – Premiere Partie – Population Légale ou De Résidence Habituelle*, Paris, Imprimerie Nationale.

- *Germany*. Rates for Prussian regions are from page 91 in Cipolla C. (1969), *Literacy and Western Development*, London, Penguin. I then took the rates of illiterate military recruits from the 1890 and 1880 *Statistisches Jahrbuch für das Deutsche Reich* and used them to linearly extrapolate the values for 1900 and 1910, as well as for the non-Prussian regions in 1870. Cipolla (1969) also used illiterate military recruits as his measure.
- *Italy*. Rates come from Felice E. (forthcoming), "Regional Convergence in Italy, 1891-2001: testing human and social capital", in *Cliometrica*, February 2012. The rates for 1910 are from page 19 in Cipolla C. (1969), *Literacy and Western Development*, London, Penguin.
- *Spain*. Nunez C-E. (1990), "Literacy and Economic Growth in Spain, 1860-1977", in G. Tortella (ed.), *Education and Economic Development since the Industrial Revolution*, Valencia, Generlitat Valencian, pp. 125-151, provides provincial literacy rates, split by gender. I took the average of this split to indicate overall literacy. Some provinces are missing from my list due to differences in aggregation. Missing regional rates are proxied with those of neighbours.
- *Sweden*. Regional rates for 1930 are from Sverige Folkrakningen 1930. These are the earliest we know of. To extrapolate back in time, I used the annual growth rate of 0.25 per cent presented in Sandbery L. and Steckel R.H. (1997), "Was Industrialisation Hazardous to your Health? Not in Sweden!", in R.H. Steckel and R. Floud (eds.), *Health and Welfare during Industrialisation*, Chicago, University of Chicago Press, pp. 127-160.

8.2 GDP Data

Table 7 below summarises the existing regional GDP data from

Caruana-Galizia and Marti-Henneberg (2013), showing the years for which data are available, number of regions, method of estimation, and currency of estimates. As mentioned earlier, most of the papers come from an international research project that aims to estimate European regional GDP in the long run, co-ordinated by Joan Roses and Nikolaus Wolf. Two sources, providing the regional GDP data for France and Germany, are independent of this project but share its aims (Schulze and Caruana-Galizia, 2013; Caruana-Galizia, 2013).

TABLE 7
Summary of regional GDP data

Country	Years	Authors	Method	Currency	Regions
Spain	1860;1900;1910	Roses et al. (2011)	Top-down	Cur. Pesetas	17
Britain	1871;1881;1891; 1901;1911	Crafts (2005)	Top-down	Cur. Pounds	12
Italy	1871;1881;1891; 1901;1911	Felice (2009)	Top-down	Con. Euros	16
A-H	1870;1880;1890; 1900;1910	Schulze (2007)	Bottom-up	Con. Dollars	22
Sweden	1870;1880;1890; 1900;1910	Enflo et al. (2010)	Top-down	Con. Kronor	24
France	1872;1886; 1901;1911	Caruana-Galizia (2013)	Top-down	Con. Francs	85
Germany	1871;1882; 1895;1907	Schulze and Caruana-Galizia (2013)	Top-down	Con. Marks	23

The benchmark years do not always match, but are never so far off to make comparisons across time unreliable. What we are after here is not precise figures of GDP at precise points in time, but approximations of relative GDP levels at intervals. This approach is more transparent than some method of interpolation or extrapolation and leads to fewer assumptions of data. In this paper, I only use the benchmark years 1870, 1900, and 1910, because they are the closest common years across all regions and because other data (literacy, mainly) were only available for these intervals.

There are two methods of estimation used in the construction of this regional GDP dataset. Top-down approaches consist of structural equations, where regional GDP levels are specified as functions of sectoral employment, wages, and value-added, as first set out by Geary and Stark (2002). Bottom-up approaches involve the painstaking application of national accounting methods to regional-level data. Does the use of different regional GDP estimation methods introduce a bias? This would matter if, for example, top-down methods consistently under- or over-estimated income levels. This is unlikely. While short-cut methods are less accurate, Geary and Stark (2002), Enflo et al. (2010), and Buyst's (2009) robustness checks against official estimates show that the margin of error is tolerable, according to national accounting standards, and that short-cuts do not produce results that veer far from officially bottom-up-constructed series. Their checks show no persistent directional bias. Furthermore, in our sample only Schulze's estimation of Austria was arrived at with a bottom-up approach. Given the above observations, in the unlikely event that these estimates are at odds with the rest of the sample, Austria only comprises 14 of the 200 regions, that is, 7 per cent of the total sample. Any bias would therefore be limited.

More interesting is the variation in currency. For our analysis, it was necessary to standardise and deflate all the estimates to make comparison across both space and time possible. We converted them into 1990 Geary-Khamis dollars ($GK), which is a standard unit in much of the literature of economics and economic history. Conveniently, Schulze's (2007) estimates for Austria-Hungary come in $GK. The estimates for Italy and Germany were already deflated; the former come in 2000 euros, the latter in 1913 marks, so their conversion was straightforward. The remaining countries – France, Britain, Spain, and Sweden –required both deflation and conversion.

The approach we took is straightforward. We deflated nominal regional GDP estimates using a national GDP deflator and then converted those estimates using the exchange rates implicit in Maddison's (2003) widely used data, which are in $GK. According to

Prados de La Escosura (2000), these data are the best of their kind available. For regional GDP estimates that were already in constant terms (Italy and Germany), we converted directly from Maddison. For consistency, we always derived the exchange rate as the period starting-year Maddison national GDP per capita divided by the starting-year national GDP per capita in our sample. For example, for Italy, which had a starting year of 1871, $ER_{Italy} = GDPpc_{Maddison, 1871} / GDPpc_{Felice, 1871}$. We then used this same exchange rate to convert all deflated regional GDP estimates. Deflators came from Smits et al. (2009), which provides datasets of nominal and real GDP, as well as GDP deflators for Spain (1958 pesetas) and Britain (1913 pounds), among other countries. France's deflator came from Toutain (1987). Enflo et al. (2010) provide both real and nominal figures in their paper.

Regional price variation is the main issue when deflating and standardising regional GDP figures. Ideally, regional price indices should be used to account for differences in prices across countries. This is rarely ever done, for three reasons. First, it is very data intensive. Very often, regional prices for a broad basket of goods defined by region simply do not exist. Indeed, Wolf writes of the difficulty of finding such data for usually well-documented Germany (Wolf 2010). This alone makes using national price indices the only way forward. Second, there is a more basic methodological concern. Apart from Schulze's (2007) data, which do not need any work anyway, the regional nominal GDP estimates are derived using the Geary-Stark method. Researchers who have adopted this method have often proxied regional wages using national wages and sometimes proxied regional wages using neighbouring regions' wages. Furthermore, these estimates are ultimately scaled to a given national GDP figure. Altogether, this makes using regional deflators inconsistent. Regional wage data often do not match and such deflators would invalidate the scaling procedure. These finer points aside, Felice (2009) argues that it is unlikely that, at least in the case of Italy, regional prices were so different that they caused differences in regional income levels. His view is supported by some of the sporadic data we have collected on regional staple goods prices. As measured by the coefficient of variation, the average regional wheat price variation in Sweden between 1870 and 1914 was just 7.62 per cent (Jorberg, 1972). In line with Felice's (2009) comments, the data underlying Jacks' (2005) work show that at least between 1870 and 1877 the variation of wheat prices between the Italian regions of Brescia, Padua, and Rome averaged 5.95 per cent. Ward and Devereux's (2003) flour price data covering 12 British cities in 1872 show that the corresponding variation was just 6.09 per cent. These levels of regional price variation are not high enough to re-order rankings of relative regional GDP per capita levels, which is the potential fundamental issue.