The Unequal Reach of Transnational Legal Institutions: Mapping, Predicting and Explaining Spatial Disparities in EU Law Use

Arthur Dyevre*, Nicolas Lampach†

Abstract

We investigate spatial disparities in the use of supranational law using a new dataset collecting the geographic coordinates of the entire universe of courts that have passed on cases to the European Court of Justice over the lifetime of the EU. We adopt a data-driven approach which integrates exploration, prediction and explanation into a single research design. Our cartography reveals that involvement in the preliminary ruling procedure tends to be concentrated in a relatively small subset of regions within member states. Next, applying machine learning methods, we find that the presence of organizations like peak courts and infrastructures like large cargo ports constitute strong predictors of EU law use. So too are capital cities and EU trademark registrations. Finally, we rationalize these findings using text mining and additional empirical tests. We argue that spatial variations are induced by the team structure of the EU court system and the division of labour within national judiciaries as well as by the resources and geographic location of litigants.

Keywords: Judicial behaviour, European Court of Justice, Machine Learning

JEL Classification: K10, K40, N34, C11, C26

* *Corresponding author:* KU Leuven Faculty of Law, Centre for Legal Theory and Empirical Jurisprudence, 45 Tiensestraat, Leuven, Belgium; E-mail: arthur.dyevre@law.kuleuven.be; Phone: +32 16325387

†, KU Leuven Faculty of Law, Centre for Legal Theory and Empirical Jurisprudence; E-mail: nicolas.lampach@kuleuven.be

Acknowledgment The authors gratefully acknowledge financial support from European Research Council Grant 638154 (EUTHORITY). We are grateful to Tomas Adamec, Gilian Bens, Frauke Petra Hein, David Ketch, Russell Neudorf and Anna Maria Tonikidou for invaluable research assistance.

1 Introduction

The European Union is the most far-reaching experiment in peaceful supranational governance in modern human history (Pinker, 2011). Its laws and regulations apply to more than 500 million Europeans, covering a wide range of policy areas, from trade in goods and services to employment, gender equality and criminal procedure. The EU as we know it today is also a fairly recent construct. The founding treaties go back to the 1950s. But multiple revisions over the last three decades have significantly altered the EU’s constitutional architecture while waves of new legislation have expanded its policy reach to new domains. Institutional change has also proceeded in less formal and more bottom-up fashion through greater coordination and cooperation between infra-state actors and EU bodies. A remarkable development, in that respect, has been the emergence of a decentralized EU legal system¹ operating on the basis of cooperation between EU courts and national courts. The EU’s implementation and enforcement capacities are extremely limited (Kelemen, 2012). So the EU must, by necessity, rely on organisations it did not create, like domestic courts, and agents it did not hire or control directly, like domestic judges and litigants, to enforce the rules it promulgates. An important gauge of the dynamics of legal integration and, by the same token, of the emergence of an EU legal system, is the number of cases passed on to the European Court of Justice (ECJ) by domestic judges.² As illustrated in Figure 1, references have risen. In marked contrast to the political process, which has been characterized by periodic stops and starts (Schneider and Cederman, 1994), with phases of rapid integration punctuated by intervals of stagnation and recurrent crises—including the empty

¹In the present paper we define the “EU legal system” or the “EU judiciary” as the the court system formed by the ECJ and national judiciaries.

²Domestic cases involving the application of EU law do not always give rise to a preliminary reference. How many such cases there are, though, remains an open question. The limited available evidence suggest that non-referred cases outnumber referred ones by one order of magnitude or more (Conant, 2002; Chalmers, 2000). Still, insofar as references are more likely to arise where judges and litigants consider EU law arguments more frequently, referral rates represent a good proxy of the intensity of EU law use.

Figure 1: Requests for a preliminary ruling passed on to the ECJ by national courts, 1961-2014.

chair face-off and, more recently, Brexit—the legal process has seemingly followed a continuously expanding trajectory. Paradoxically, at a time when the legitimacy of EU policy making is contested, sometimes vehemently, the ECJ—which, from its seat in Luxembourg, presides over the operations of the EU’s decentralized judiciary—is viewed as more powerful than ever. Easily ranking among the world’s mightiest judicial bodies, the Court has been listed by British Prime Minister Theresa May as one of the principal justifications for Brexit:

[W]e will take back control of our laws and bring an end to the jurisdiction of the European Court of Justice in Britain. Leaving the European Union will mean that our laws will be made in Westminster, Edinburgh, Cardiff and Belfast. And those laws will be interpreted by judges not in Luxembourg but in courts across this country.³

The emergence and expansion of the EU judiciary raise important questions about the effectiveness of governance structures relying on decentralized enforcement. First, one may question

³Theresa May, 17 January 2017.

the incentives and constraints of the domestic agents enlisted in the enforcement effort. How committed should we expect them to be? Domestic judges called on to enforce EU regulations are not on the payroll of the EU and may consider themselves busy enough already with the task of applying national law. All the more so as there are few levers EU courts can turn to to coax their domestic counterparts into compliance. Unlike national judiciaries, the EU judicial system is not hierarchically structured. EU courts do not hold the power to reverse domestic rulings. In turn, the fear that, due to the judges' ignorance or wilful disregard, EU law arguments could lack traction in the courtroom may make litigants wary of pursuing an EU law strategy. Second, both the agents and the domestic context in which they operate differ considerably, not only across but also within member states. Should we expect these difference to impact the application of supranational law? While top court judges can draw on greater resources to apply what may otherwise form a less familiar body of law, they may also resent EU judges encroaching on their sphere of influence. Courts at the lower echelons of the domestic judicial hierarchy, on the other hand, typically have less resources to allocate to a comparatively larger caseload, potentially raising the opportunity cost of submitting references. Furthermore, in less prosperous parts of the EU territory, courts as well as litigants may be reluctant to embrace legal rules whose application appears more resource-intensive. Should we expect EU law to have the same impact in poor, rural Algarve as in wealthy, urbanized Vienna? Finally, we may wonder whether decentralised modes of governance and nonhierarchical judiciaries are subject to intrinsic limitations. Should we expect cross-court and cross-regional variations to diminish or, on the contrary, to become more pronounced over time? Institutional diffusion may induce more judges to embrace legal integration. Yet lingering uncertainty over judicial enforcement may make EU litigation the preserve of litigants with deep pockets who can expect high gains despite low odds of success or can afford forum shopping.

Although their formulation may differ, these puzzles have received considerable attention in the large interdisciplinary literature investigating the European integration process. They have been the focus of significant theoretical and empirical efforts by scholars variously hailing from law ([Weiler, 1991](#)), history ([Rasmussen, 2012](#); [Davies and Rasmussen, 2012](#); [Davies, 2012](#)), comparative politics ([Sweet and Brunell, 1998](#); [Conant, 2002](#)), sociology ([Cohen and Vauchez, 2007](#); [Vauchez, 2010](#); [Fligstein and Stone Sweet, 2002](#)), international relations ([Garrett, 1995](#); [Alter, 2001](#)) and economics ([Pitarakis and Tridimas, 2003](#)). While the EU is, in many ways, a unique experiment, the questions raised by its governance structure transcend the European context. They are fully relevant to the rising number of international legal regimes that have sought to emulate the EU ([Alter, 2012](#); [Alter and Helfer, 2010](#)). But they also relate to debates on the connection between economic development and access to litigation within states ([Eisenberg et al., 2012](#); [Jacobi, 2009a](#)) and economic inequality in litigation activity ([Galanter, 1974](#); [Songer et al., 1999](#)). Yet, despite the vastness of this literature, our understanding of legal integration appears both fragile and fragmented. Several studies have attempted to identify the determinants of referral activity in the EU ([Sweet and Brunell, 1998](#); [Pitarakis and Tridimas, 2003](#); [Wind et al., 2009](#); [Hornuf and Voigt, 2015](#); [Carrubba and Murrah, 2005](#); [Fligstein and Stone Sweet, 2002](#); [Vink et al., 2009](#); [Dyevre and Lampach, 2017](#)). These studies, though, reveal serious weaknesses. Nearly all of them rely on data aggregated at country level. For that reason, they cannot disentangle the separate effects of adjudication and litigation dynamics, although this is arguably essential to our understanding of legal integration. Research has been further hampered by the relative dearth or incompleteness of data on national judiciaries, law firms and litigiousness.⁴ As for the more qualitative studies that have attempted to disaggregate the data to look at patterns of EU law use within states, they have rarely considered more than two or

⁴Existing cross-country datasets on litigation and judicial institutions, such as CEPEJ, suffer from multiple limitations, including restricted temporal coverage and level of aggregation (court-tier rather than court or region).

three member states at a time (Kelemen, 2011; Alter, 2001; Conant, 2002; Kelemen and Pavone, 2016, 2018; Chalmers, 2000; Alter and Vargas, 2000; Cichowski, 2007). Though this strand of scholarship has been better able to track and document the behaviour of domestic judges and litigants, the nature of their research design implies that their findings cannot be generalized to the rest of the EU.

We attempt to remedy some of the limitations of this literature by proposing a novel empirical approach that integrates exploration, prediction and explanation into a single research design. Similar to Kelemen and Pavone (2016) and Kelemen and Pavone (2018), we focus on sub-national dynamics. This allows us to exploit variations in the location of litigants and judicial institutions to better untangle their respective impact on referral rates. We present a new dataset compiling the geographic coordinates of the entire universe of courts that have submitted a request for a preliminary ruling up to 2014. We use this data to map the geography of referral activity. Our cartography of EU law use highlights important trends and patterns. First, involvement in the preliminary ruling procedure tends to be concentrated in a relatively small subset of regions within member states, although some member states, including Italy and Germany, display lower levels of spatial concentration. Regions that are home to the country’s capital typically constitute the main cluster of referral activity. Second, over time, top courts have come to account for an increasing share of submissions, to the point that they now dominate the preliminary ruling procedure. While maps and other data visualizing techniques are proven exploratory tools, we use machine learning methods to further explore referral dynamics and to identify predictors and promising explanatory hypotheses. To that end, we assign the courts to the 273 NUTS-2 regions comprising the EU. Doing so permits us to relate regional referral rates to both national and regional attributes. Using Conditional Inference Tree (CIT) in combination with Random Forest (RF), we explore a broad range of potential predictors of spatial variations and the

relations among them over the period 2000-2014 (at the time of writing). Besides demonstrating the advantages of machine learning methods over more conventional regression techniques, our approach shows that peak court, capital city, EU trademark registrations, GDP and cargo ports constitute strong predictors of referral activity. Furthermore, our machine learning methodology enables us to identify possible interaction effects, such as between peak court and economic prosperity. Our approach is inductive in the sense that we deliberately let the data speak first with a minimum of assumptions. Only in the final, explanatory stage of our inquiry do we attempt to rationalize the patterns uncovered by our machine learning models. Our data-driven rationalization draws on team models of adjudication and economic explanations of litigation behaviour a discussion. We also apply text-mining tools and additional empirical tests to refine our interpretation of the predictive and exploratory part of the study. We argue that spatial and temporal variations in referral activity are induced by the structure of the EU legal system as well as by the incentives of judges and litigants. First, the referral procedure put in place by Article 267 of the Rome Treaty structures inter-court relations according to a pure team model. Domestic courts self select to cooperate with ECJ and thus operate as gatekeepers of the procedure. In what is effectively a nonhierarchical legal system, the EU apex court cannot rely on litigant selection to control the behaviour of recalcitrant judicial agents. Second, referral rates are affected by the division of labour and resources within national judiciaries. The peak courts' focus on law-finding and law creation together with the self-reinforcing institutionalization of Article 267 have turned them into dominant players of the referral procedure. Third, we suggest that litigants representing trade interests and the wealthy have stronger incentives to invoke EU law. A textual comparison of the cases coming from port regions with the cases coming from other regions validates the hypothesis that referral activity in port regions is driven by their privileged position as trade hub. EU trademark registrations, we contend, are similarly related

to patterns of trade. The association between regional GDP and referral activity in regions hosting peak courts further points to a link between EU law litigation and economic prosperity. While there are many questions that our study leaves to future research, we contend that the institutional economic narrative of EU law emerging from our analysis offers a strong baseline from which to further investigate the behaviour of judges and litigants. With future research in mind, we specify an alternative, parsimonious regression model on the basis of the insights gained from our inductive approach. This alternative baseline model generalizes well to test data while its application to the smaller data set compiled by [Kelemen and Pavone \(2018\)](#) casts doubt on the effect of specialized law firms on referral rates.

2 Relationship to the Literature

Our approach and substantive analysis relate to several literatures. Both economics and political science have seen a surge of interest in the geographic dimension of economic and political behaviour. As part of a strain of research described as “New Economic Geography”, economists have investigated the effect of development and trade liberalization on economic disparities within countries ([Lessmann, 2014](#); [Brühlhart, 2011](#); [Henderson, 2000](#); [Brühlhart, 2001](#); [Brühlhart and Koenig, 2006](#); [Rodríguez-Pose and Gill, 2006](#)). This work explicitly considers the impact of European integration on spatial inequalities ([Brühlhart, 2001](#); [Rodríguez-Pose and Ezcurra, 2009](#); [Brühlhart and Koenig, 2006](#); [Melchior, 2012](#)). Some of our findings, notably the high level of referral activity in port regions and capital cities, can be directly related to this literature.

Like economists, political scientists and empirical legal scholars are paying greater attention to geography. Maps have become a more frequent feature in journal publications as researchers have tried to explain infra-state spatial variations in political behaviour ([Howe et al., 2015](#); [Hersh and](#)

Nall, 2016; Franzese Jr and Hays, 2008). This evolution has spilled over to the study of courts and litigation. Eisenberg et al. (2012) and Jacobi (2009a) consider variations in litigation rates across states in, respectively, India and the United States. Both papers report a positive association between economic prosperity and litigation rates. Examining court participation in Article 267 proceedings across Italian regions, Kelemen and Pavone (2016) directly relates to the present study. Litigiousness is found to be associated with greater referral activity in northern Italian regions but not in southern regions (Kelemen and Pavone, 2016). In a companion paper (Kelemen and Pavone, 2018), Pavone and Kelemen extend their analysis to France and Germany. They find referral activity to be more centralized in France than in Germany and Italy, a difference they link to variations in the institutional structure of the judiciary across the three countries. All three countries, though, display a trend towards greater involvement of apex courts in referral proceedings. Their work further report strong referral activity in regions with more specialized law firms as well as a higher incidence of free movement references in the port regions and manufacturing hubs of Germany and northern Italy (Kelemen and Pavone, 2018). While also adopting a geographic approach, our study is less concerned with the temporal sequence of EU law diffusion and more with the determinants of referral activity. Our work considers all member states and investigates a much broader range of possible predictors of referral activity, although missing observations for many covariates forces to restrict our analysis to a shorter time frame.

The work of Kelemen and Pavone forms part of a large empirical literature on the preliminary ruling procedure, which encompasses qualitative as well as quantitative work. Quantitative multi-country analyses have sought to identify determinants of referral propensity using country-level data. Owing to the difficulty inherent in attempting to infer micro-level trends from macro-level data, though, these efforts have met with mixed success. In particular, such data does not allow researchers to untangle the distinct influences of litigation dynamics and judicial behaviour. This,

besides methodological limitations, may explain the conflicting conclusions of this scholarship, whether it is over the effect of intra-EU trade, monism or public support for EU membership (Sweet and Brunell, 1998; Pitarakis and Tridimas, 2003; Carrubba and Murrah, 2005; Hornuf and Voigt, 2015; Wind et al., 2009; Vink et al., 2009). Applying both multi-level and machine learning modelling, Dyevre and Lampach (2018) suggest that only trade, population and years of EU membership constitute reliable, country-level predictors of referral frequency. Of those three, only trade and membership duration seem to offer insights of some sort into the beliefs and motives of litigants and judges. Compared to country-level measurements, regional-level data permits a more fine-grained analysis of litigation and adjudication dynamics. Courts, economic hubs, political institutions and big cities are located in different regions, providing variations whose effect cannot be investigated at country level but which our study is able to exploit.

In a more qualitative vein, Conant (2002) emphasises the limits of legal integration and argues that the narrow base and high access costs of EU law limits its diffusion. Our findings are consonant with this narrative which we relate to the New Economic Geography literature. On the other hand, whereas many EU scholars stressed the role of lower courts in the enforcement of EU law (Alter, 2001; Weiler, 1991), our data shows that the preliminary ruling procedure is now dominated by supreme courts. We relate this evolution to the literature on institutional change (North, 1990; Kingston and Caballero, 2009; Greif and Kingston, 2011; Zweynert, 2009; Brousseau et al., 2011; Brousseau and Raynaud, 2011) and suggest changes in informal practices interacted with existing formal rules have induced shifts in the referral behaviour of higher court judges.

Our study also relates to theoretical work on judicial institutions and the nature of inter-court relations. This work has emphasised two distinct rationales for inter-court cooperation. Agency models emphasise interest divergence and focus on the mechanisms—strategic auditing,

litigant selection, judicial whistle-blowing—enabling higher courts to control the behaviour of their judicial subordinates ([Cameron et al., 2000](#); [Beim et al., 2014](#); [Kastellec, 2016](#)). Team models, meanwhile, posit that judges share a common goal, which is to maximize the number of “correct” adjudicatory outcomes given available resources and informational constraints. Accordingly, courts cooperate across layers of the judiciary not because they anticipate reversal but because specialisation of labour makes this goal easier to achieve ([Kornhauser, 1994](#); [Kastellec, 2016](#); [Cameron and Kornhauser, 2006](#)). We argue that the EU legal system is better characterised as conforming to a team rather than an agency model. At least below the level of apex courts, the decision to request a preliminary ruling presupposes that the domestic court and the ECJ share a common goal (or least that the domestic court believes so) and that dispute resolution will benefit from the ECJ’s input. The referral mechanism makes agency problems such as non-compliance difficult to tackle, inducing greater geographic variations in EU law use than would be the case under an appeal mechanism. Similar to [Romeu \(2006\)](#), we also argue that differences in fact-finding vs law-finding specialisation, resources and caseload give domestic peak courts greater incentives to use EU law.

Finally, our study contributes to the expanding literature on applied predictive modelling in economics and political science ([Varian, 2014](#); [Cranmer and Desmarais, 2017](#)). Machine learning and statistics are often seen as separate disciplines pursuing separate goals, with machine learning emphasising prediction and statistics hypothesis-testing ([Breiman et al., 2001](#)). Like [Varian \(2014\)](#) and [Cranmer and Desmarais \(2017\)](#), we argue that predictive and machine learning models do not only represent a powerful exploratory tool but can also greatly enhance theory-building, theory choice and theory-testing. Specifically, we believe that this applies with particular force when the availability of a new dataset offers researchers the opportunity to reconsider the empirical assumptions of a specific field of inquiry.

3 Prediction, Exploration and Explanation

This Section outlines our inductive research design. We begin by explaining how, in some contexts, a more inductive, data-driven approach can remedy the pitfalls of what we call the “conventional social science method”. We make the case for machine learning as a tool to integrate prediction, exploration and explanation. In so doing, we are careful to stress that data-driven research should not be equated with theory-free or assumption-free research. Theory is still needed, although less so at the prediction and exploration stage than at the explanatory-or-validation-stage of the research.

3.1 The Conventional Social Science Method

Social scientists commonly view prediction, exploration and explanation as distinct endeavours calling for separate modelling approaches. In disciplines such as economics and political science, the emphasis has traditionally been on explanation. This has shaped not only the way researchers approach and model data, but also the way they report quantitative findings ([Cramer and Desmarais, 2017](#); [Varian, 2014](#)). The typical, textbook approach to quantitative social science research requires that the researcher specifies a statistical model on the basis of theory, distributional fit and, in some instances (e.g. panel or clustered data), dependencies assumed to exist in the data structure. In order to work, the traditional approach presumes that the researcher already knows a lot a priori, that is before the first datapoint is crunched. What functional form captures the relationship between covariates and outcome? What controls should be included in the analysis? What interaction effect, if any, should be taken into account? How does the nested or panel structure of the data influence the subjects’ response to the covariates? In practice, explanatory modelling is rarely as straightforward as textbook examples or journal

articles frequently suggest. One reason is that theoretical intuitions tend to lack the level of detail required to specify a functional form. As a result, researchers often end up specifying an additive model—that is, one in which the response variable is expressed as a linear combination of covariates—which they then estimate using regression analysis. In the many instances where this fails to produce the expected outcome, researchers respond by tinkering with the initial model, adding or removing covariates and interaction terms before running a new estimation. It is not uncommon, especially when confronted with a new data set and/or when the number of candidate regressors is large, for researchers to iterate this procedure over multiple rounds (Sala-i Martin, 1997; Hendry and Krolzig, 2004). So dozens, sometimes hundreds, of regressions are run until a result deemed worth reporting is eventually reached.

This suggests that, in practice, explanatory modelling commonly involves data exploration as a prior, albeit unreported, step. Widespread as it is, though, this *modus operandi* is neither an efficient nor a transparent way to explore data, irrespective of whether the ultimate goal is explanation or prediction. Patterns present in the data but which are not suggested by available theories can easily be overlooked. A bad problem is made worse by the use of regression analysis. Regression techniques are not well-suited to handle models with large numbers of predictors (Dormann et al., 2013). Yet kitchen sink regression models are ubiquitous. In the Article 267 literature, for example, authors commonly report the estimates of models with 10 to 15 covariates (see e.g. Wind et al., 2009; Hornuf and Voigt, 2015; Carrubba and Murrah, 2005). Nor does the reliance on in-sample measures of fit do much to help the cause of empirical researchers. Good in-sample fit is no guarantee that a model will generalize well to new data. Over-emphasis on in-sample statistical significance can easily make one oblivious to the danger of overfitting. Again, taking the preliminary reference literature as illustration, we observe that no paper reports any measure of out-of-sample accuracy, whether RMSE or out-of-sample R^2 . So we may have

legitimate suspicion that some of these models fit the sample used to estimate them too closely.

Such practices can greatly hamper empirical as well as theoretical progress. As published results obtained from overfitting kitchen sink regressions come to define which controls subsequent analyses are expected to include in their model specification, they can create a perverse ecosystem of studies, giving rise to a vicious circle of models capturing ever-greater statistical noise. Not implausibly, this may reflect the current state of quantitative scholarship on the preliminary ruling mechanism with its long list of contradictory findings.

Methodologists have discussed possible cures for these pathologies. An old, but proven recommendation is to plot the data. For problems with few variables, the ability of the human eye to spot patterns is unmatched. Upholding this time-honoured precept of statistical wisdom, our cartography of referral activity shows that visualisations represent indeed a precious tool to uncover and explore relationships in data. Besides graphing the data, some authors have advocated the use of cross-tabulation along with a drastic reduction in the number of covariates included in model specifications ([Achen, 2002](#)). Yet these fixes have limitations of their own. Researchers also care about omitted variable bias while plots and cross-tabulation make data exploration impossibly cumbersome when the number of covariates exceeds ten or more.

3.2 Machine Learning and Prediction as Criterion for Hypothesis Selection

Machine learning methods potentially offer a solution to these issues. First, instead of assuming that the analyst knows a priori what patterns exist in the data, machine learning methods allow him to approach the data with greater humility, acknowledging rather than concealing his relative ignorance. Keeping assumptions to a minimum, the researcher can let the data speak before deciding which covariates matter and could represent a plausible hypothesis. Many machine learning methods, including the ones deployed in the present study, are non-parametric.

So they do not require the data analyst to make distributional assumptions. In that sense machine learning methods can be compared to exploratory plots, except that they scale up to larger and more complex exploratory tasks. Second, machine learning techniques, especially single-tree methods, can help researchers spot interaction effects or non-monotonic effects among predictor variables (Varian, 2014). Third, machine learning methods can handle large numbers of covariates (Dormann et al., 2013; Fokkema et al., 2017), which comes especially handy in the context of the present study, as we seek to assess the effect of a wide array of factors, some of which have not yet been considered in the literature. Finally, predictive power provides a solid criterion to identify the most promising set of hypotheses or theories while averting overfitting. Generally speaking, there should be no dispute that, other things being equal, research should prioritize the hypotheses that have the strongest predictive power.

Let f^* be the function that maps input data into the output of interest. The traditional, prescribed *modus operandi* of empirical social science research presupposes that the researcher knows from theory the function $\hat{f}$ that approximates f^* . In contrast, machine learning seeks to approximate, or “learn”, the function $\hat{f}$ that best captures f^* from a class of functions F by using the data itself. This involves the following steps. First, partition the data into training and test sets. Then fit all functions $f \in F$ to the training set. Finally, measure their performances against the test set. The function $\hat{f}$ selected as the best approximation of f^* is the function $f \in F$ that performs best against the test set. In practice, in the case of a numeric dependent variable, the measure of out-of-sample accuracy is RMSE or R^2 . The two machine learning algorithms we apply in this paper, Conditional Inference Tree (CIT) and Random Forest, follow the same basic procedure. Cross-validation ensures that $\hat{f}$ effectively generalizes to new data, thereby averting overfitting.

To be sure, because $\hat{f}$ is generated from the data rather than derived from theory, there is

no absolute guarantee that it identifies meaningful relationships. For that reason, unless one is solely interested in prediction, the results of machine learning methods need ex post validation. This is where theory and more conventional statistical methods have their full, legitimate place.

3.3 Explanation as Data-driven Rationalization

Under the approach adopted here, the last, explanatory, stage of research can be characterised as a rationalization exercise. How to make sense of the patterns identified in the data? What causal inference can we draw? Method-wise, there is no one-size-fits-all answer to this question. For our part, we use a combination of theory, text-mining and statistical testing (notably fixed-effects panel regression). That aspect of our study, at least, is more in line with the conventional social science model.

4 Mapping Spatial Disparities in Referral Activity

To construct our outcome variable and explore spatial variations in EU law use we collected the geographic coordinates of all the courts that submitted at least one preliminary reference over the period 1961-2014 from Google maps. For each of the more than 8000 references issued by national courts, the following information was also coded either manually or by scraping the data directly from the ECJ Curia website: member state of origin; case number; date on which the Court of Justice received the reference; date on which the Court of Justice issued its decision (whether order or ruling); name of the referring court; and position of the referring court in the domestic judicial hierarchy (first instance, intermediate, peak court).⁵

As said above, maps, along with other visualization techniques, represent a proven and powerful method to explore data. Depicted in Figure 2 are cumulated referrals across the EU territory

⁵We were able to check some of that information against a dataset compiled by [Sweet and Brunell \(1998\)](#).

for the period 1961-1972, 1973-1985, 1986-1994, 1995-2003, 2004-2014 and 1961-2014.

(a) 1961-1972

(b) 1973-1985

Figure 2: Referral activity across EU territory. Colour denotes court level in domestic judicial hierarchy. Dot size represents number of referrals. Panel (a) plots referral activity in EU-6 (Belgium, France, Germany, Italy, Netherlands and Luxembourg). Panel (b) plots referral activity in EU-10 (UK, Denmark and Ireland joined in 1973, Greece in 1981).

Figure 3: Referral activity across EU territory. Panel (a) plots referral activity for EU-12 (Spain and Portugal joined in 1986). So does panel (b) for EU-15 (Austria, Finland and Sweden joined in 1995).

Figure 4: Referral activity across EU territory. Panel (a) shows referral activity in EU-28 (13 countries acceded to the EU in three waves: 2004, 2007 and 2013). Panel (b) shows cumulated referral activity over the lifetime of the EU.

Striking is, first, the level of concentration of referral activity. With the exception of Italy, Germany and Romania, participation in the preliminary ruling procedure tends to be highly concentrated in one region, in most cases the capital, as shown in Figure 5. Most references from the UK have been submitted by courts in London, just as Greek references have tended to come from Athens, Danish ones from Copenhagen and Irish ones from Dublin. In that sense, member states, even federal ones such as Austria, look in their majority more like France and its allegedly centralized judiciary (Kelemen and Pavone, 2018) than like Germany and Italy. The Netherlands is a special—and somewhat misleading—case. Officially, Amsterdam, in the province of Northern Holland, is the country’s capital. Yet the seat of the government, together with that of the two main Dutch supreme courts, is in the Hague, in the province of Southern Holland, which, on top of accounting for the largest share of Dutch references, exhibit the highest referral activity of all EU regions.

Figure 5: Share of references coming from courts in the capital region (NUTS-2), 2000-2014. Dashed line shows share excluding references from apex courts. For the UK figures are aggregated from the five NUTS-2 regions comprised in Greater London.

Second, through time peak courts have become larger purveyors of preliminary references, as evidenced by the arc of large reddish dots going from The Hague (where Dutch supreme courts are located) to Helsinki (seat of the two Finnish supreme courts) for the period 2004-2014. Exceptions include the UK, where intermediate courts dominate proceedings, along with Hungary and Bulgaria, where supreme and constitutional court judges have made very limited use of Article 267. Top courts still account for the largest dots in the three countries where none is located in the capital, namely the Czech Republic, where peak courts are all located in Brno, Estonia, where Tartu (not Tallinn) is the seat of the Supreme Court, and Germany, where peak courts are located in five different cities.⁶ Kelemen and Pavone (2018) had already detected

⁶The *Bundesverwaltungsgericht* initially had its seat in West Berlin, but the Court was relocated to Leipzig in 1997, three years after the capital of reunified Germany was transferred from Bonn to Berlin.

a shift towards greater involvement of higher courts in France, Italy and Germany. As shown clearly in Figure 6, references from first instance courts were more numerous until 2000 but also more volatile. Requests from peak and intermediate courts, by contrast, initially grew at a slower but steadier pace until peak court overtook lower courts at the turn of the millennium.

Figure 6: Referral activity by court level, 1961-2014.

Beyond capital cities and apex courts, the maps indicate that large cities and economic hubs such as Barcelona, Milan, Lyon, Hamburg and Düsseldorf produce a steady stream of references.

Overall, our cartography of Article 267 use suggests that referral activity tends to cluster in regions that enjoy a particular status either as seat of the country’s top courts or of its capital or as economic hub. Outside these clusters referral activity exhibit greater volatility, in part because lower courts are less likely to be repeat players in the preliminary ruling system.

Aside from the presence of a peak court or hosting the country’s capital, there are many national and regional factors which we may think of as potentially having an influence on referral rates. As the number of potential factors reaches 20 or more, visualization tools become inefficient. Hence our recourse to machine learning techniques.

5 Identifying Predictors of Referral Activity

This Section presents our machine learning procedure together with the data used for predicting references. After presenting the data and unit of territorial analysis, we introduce the two machine learning algorithms employed in our analysis, Random Forest and Conditional Inference Tree. Finally, we present the results of our machine learning experiment, which we compare to a baseline, additive regression model.

5.1 Territorial Unit of Analysis

To explore the factors affecting regional variations in referral activity, we begin by assigning the referring courts to the NUTS-2 territorial unit ⁷ in which they are located using Google Maps data. NUTS-2 regions correspond to Spanish “*Comunidades*”, French “*régions*”, German “*Regierungsbezirken*” or Belgian “*provinces*”. Adding Croatia, which joined the EU in 2013, the total adds up to 273 regions.

5.2 Covariates

While requiring less a priori assumptions, inductive designs can never be entirely assumption-free. Save for purely predictive tasks, this also holds for the selection of variables included in the analysis. Data-driven approaches do not require one to make a strong theoretical case to justify the inclusion of a particular covariate in the analysis. Instead, a more lenient “plausibility” criterion can be applied. However, while a looser criterion can help reveal unexpected insights, it may also make the interpretation and rationalization of the results more difficult. In that sense, there is an inclusiveness-interpretability trade-off. In our case, we include, first,

⁷“NUTS”, which stands for Classification of Territorial Units for Statistics (or “*Nomenclature des unités territoriales statistiques*” in French), is a geocode standard developed and used by EU institutions. Our choice of NUTS-2 over NUTS-3 is motivated by the availability of covariates for NUTS-2 as well as by the difficulty of matching geographic coordinates—which we found to be considerably more difficult for the smaller NUTS-3 units.

all the covariates that have been tested in the Article 267 literature for which data is available for all member states. Second, we include variables that have been suggested but not tested by prior studies, such as the presence of a cargo port in the territorial unit ([Kelemen and Pavone, 2018](#)). Additionally, we include national and regional-level variables that could plausibly have an effect on referral activity such as rule of law indicators, EU trademark registrations, employment, primary income, urbanization and migration. Finally, we add a couple of less obvious regional indicators such as the number of vehicles and the number of victims of car accidents. The complete list of covariates together with their source, definition and operationalization, is reported in the Supplementary Materials. The list encompasses legal, economic, demographic, environment and governance attributes measured either at the national or at the regional level. We also included the regions' name and the country in which they are located as additional information in our analysis. As reported in the Supplementary Materials, some of the variables included are highly correlated. This applies in particular for economic (GDP, primary income, employment, trademark registrations) and governance (government efficiency, rule of law, voice, political stability) indicators. Unfortunately, cross-country or cross-regional data on litigiousness is very scarce, although urbanization along with various measures of economic development have been found to be positively associated with litigation rates ([Jacobi, 2009b](#); [Eisenberg et al., 2012](#)). Owing to limitations in data availability, we restrict our primary analysis to the period 2000-2014. We fitted, tested and selected our machine learning models using complete cases, but we also report results using predictive mean matching (PMM) to impute missing data in the Supplementary Materials.

Some covariates, such as the presence of a peak court or a monist tradition, relate more to the judicial side of legal integration whereas indicators of economic prosperity or trade activity, like the presence of a large cargo port, are more likely to capture litigation dynamics. With

respect to the judicial dimension of the argument, it is important to bear in mind the limitations of our data. While it allows us to locate all peak courts over the lifetime of the EU, it does not include information on the intermediate and first instance courts that have never referred cases to the ECJ. Although we can safely assume that there are courts in all regions, we cannot say how many or what the exact proportion of first instance and intermediate courts is. However, despite these limitations, what we know about the location of peak courts does permit us to disentangle, at least to some extent, the impact of judicial variables on referral activity.

5.3 Machine Learning Models: Random Forest and Conditional Inference Tree

Machine learning methods are often applied to purely predictive tasks. This justifies the use of methods allowing arbitrarily complex F such as neural networks and ensemble methods. While these methods possess great predictive accuracy, they yield a $\hat{f}$ that is essentially a black box. Here our goal is not purely predictive. Although we do want to know which covariates are strong predictors of regional variations, we are also interested in gaining insights into the data generating process, which presupposes an interpretable $\hat{f}$. In the context of such an inquiry, Conditional Inference Tree (CIT) ([Hothorn et al., 2006](#)) and Random Forest (RF) ([Breiman, 2001](#)) possess complementary strengths. CIT is a decision learning method similar to Classification and Regression Tree (CART). As with CART, it defines disjoint subsets of the data–population subgroups—through recursive binary partitioning of the vector space. Partitions are defined so as to minimize the discrepancy between predicted and observed values of the outcome variable. The sequence of partitions can be represented as a single tree, where nodes indicate the covariate values used to split the sample into subgroups and leafs (terminal nodes) indicate the distribution of the values of the outcome variable in the resultant subgroups. Single-tree methods like CART

and CIT yield a readily interpretable $\hat{f}$. The size of the tree directly relates to the precision-generalization trade-off, with larger trees resulting in more precise predictions but also in larger generalization error. The shape of the tree, meanwhile, can be interpreted as the combination of attributes associated with each subgroup. Single-tree methods are particularly helpful to identify non-linear patterns and interaction effects and can provide direct insights into the underlying data generating mechanism. The main difference between CART and CIT is that the latter adds a non-parametric significance criterion for both variable selection and stopping. This remedies CART’s selection bias towards covariates with many splits or missing values and result in more reliable trees (Hothorn et al., 2006).

The principal limitation of single-tree methods is that they can be unstable, due to the tendency of single trees to overfit their training set (Hayes et al., 2015), although this problem is less pronounced in the case of CIT and partly remedied by cross-validation (Hothorn et al., 2006). As it has been specifically designed to prevent overfitting, Random Forest (RF) is a useful complement to CIT. RF is one of the most popular machine learning algorithms (Breiman, 2001; Heidema et al., 2006). Like CIT, it operates by constructing decision trees. However, instead of growing a single tree, it grows a multitude of them by repeatedly selecting random samples (with replacement) of the training set and fitting them new trees. To grow a tree, the algorithm selects, at each candidate split, a random subset of the covariates. It then takes the average prediction of all the trees in the forest thus grown. RF produces less readily interpretable results—forests comprising thousands of trees are hard to visualize—but also smaller generalization error. It thus provides researcher with a highly reliable measure of the ability of a set of variables to predict the outcome of interest. In the present study, we essentially use RF as a consistency check for the more readily interpretable CIT results.

As for any machine learning procedure, finding $\hat{f} \in F$ amounts to finding the function $f \in F$

with the optimal parameter value(s), that is the parameter values that minimize the discrepancy between predicted and observed outcomes in the test data. In computer science, this process is known as “parameter tuning”. Here the relevant tuning parameters are the minimum significance criterion p in the case of CIT (Hothorn et al., 2006) and the maximum number of trees and the number of variables to randomly select at each node split for RF (Breiman, 2001). We define the tuning grid as $0.950 \geq 1 - p \geq 0.999$ for the minimum significance criterion and as $1 \geq N \geq \max$ for the number of variables to be randomly selected. We set the minimum number of trees at 1000—which is common practice for RF models. Parameter tuning occurs via cross-validation, that is by comparing how f fitted on the training data perform on the test data. For both RF and CIT, we employ 10-fold cross-validations to train and test the models and repeat this procedure five times. Models are evaluated on the basis of their root mean square error (RMSE). In addition, we apply the one-SE rule: we select the simplest model that is one standard deviation from the model with the lowest RMSE value.

We compare RF and CIT to a baseline frequentist, multi-level regression model with with level-1 equation specified as:

$$Y_{ijt} = \exp(\beta_{0jt} + \beta_{1jt}X_{ijt} + e_{ijt}) \quad (1)$$

$$(2)$$

where Y_{ijt} is the number of references in region i in member state j in year t , X_{ijt} refers to the vector of region-level predictors and X_{jt} to the vector of country-level predictors. In order to facilitate comparison with the results yielded by machine learning, we use the same 10-fold cross-validation procedure to fit the model and compute the RMSE.

5.4 Results

Tuning plots are reported in the Supplementary Materials. The models presented here do not include country-level intra-EU trade as predictor. Because trade data is entirely missing for some countries, its inclusion entails a drastic reduction in the number of complete observations. We report results for models including this covariate in the Supplementary Materials. Intra-EU trade is found to have little predictive power.

Figure 7 compares the performances of the baseline regression model with RF and CIT. The two machine learning methods consistently outperform the baseline regression model, producing significantly smaller RMSE and displaying significantly less variance in their out-of-sample performances. The RF model displays the highest level of accuracy with $\text{RMSE} = 1.5$ and $R^2=0.80$. As expected, CIT does less well than RF, though still significantly better than the regression model.

Figure 7: Comparing performances of baseline multi-level regression, CIT and RF.

Looking at the contribution of variables to prediction, we see that *peak court* comes out as the strongest predictor in both the RF and CIT model (Figure 8). As with our cartography of referral activity, this suggests that the particular incentives of peak court judges to submit

references should be an important part of a theory of legal integration. Other strong predictors include capital, GDP, cargo port, EU trademark registrations, urbanization, primary income, employment, population and vehicles. While cargo port may intuitively seem to reflect the effect of trade hubs on litigation, GDP, trademark registrations, primary income, employment and vehicles are conventionally interpreted as measures of wealth and economic dynamism. “Region” and “Country” in the CIT model and the regional IDs in the RF model point to the effect of country and region-specific patterns. According to the RF model, Tyrol, and to a lesser extent Vienna and Southern Holland, produce more references than other variables otherwise predict. For comparison, Figure 9 illustrates the coefficient values for the variables included in our baseline regression model. A look at the confidence intervals that do not include zero reveals that the model finds the variables *peak court*, *population*, *cargo port*, *government efficiency*, *judicial review*, *constitutional court* and *capital* to be statistically significant. These also score high in the machine learning models, except, interestingly, for *government efficiency*, *judicial review* and *constitutional court*.

Figure 8: Importance of variables to prediction. Performance metric is the reduction in node impurity.

Figure 9: Baseline multilevel regression model. Plot shows regression coefficients for all the variables included in models computed via 10-fold cross-validation and fit on the whole dataset. ‘*’ denotes significant predictors.

Besides helping to discern strong predictors, machine learning methods can assist in identifying interaction and non-monotonic effects. As said before, this is particularly true of single-tree procedures like CIT. Depicted in Figure 10 is the tree generated by our optimized CIT.⁸

Figure 10: Conditional inference tree. Tree is generated using all the variables in Table ?? along with regional IDs. Significance criterion is $1 - p = 0.99$.

CIT techniques are well-suited for the analysis of clustering effects, where, as here, subpopulations have different mean values. Here the tree suggests two causal pathways determined by the presence or absence of a peak court in the region. If the region is the seat of a peak court, GDP per capita predicts regional divergence, with GDP above EUR 60 636 associated with considerably higher referral rates. If there is no peak court in the region, then a differential in EU trademark registrations can lead to divergence in referral dynamics, although a lower level of EU trademark registrations can be partly offset by the presence of a cargo port or the country’s capital city. Regions for which these attributes are absent—the overwhelming majority

⁸Plot for the tree grown without regional IDs is included in the Supplementary Materials.

of observations in our sample—show referral rates that rarely deviate from zero. Intuitively, the variables in the tree may be related to the twin influences of litigation and adjudication dynamics. Peak court points to the role of variations in judicial structures, while the other variables point to the resources and location of litigants with incentives to invoke EU law.

That variables like employment and urbanization do not appear in the tree may come as a surprise considering their strong predictive power (see Figure 8). This can be readily explained by CIT focusing on the attributes that are most distinctive of the subpopulations. Urbanisation, for example, is associated with higher referral rates (so regions in node 5 tend to be more rural) but its level varies little among nodes 6, 7, 8, 10 and 11.⁹ In the next Section, we move on to the explanatory stage of our study in an attempt to rationalize these results. We concentrate on the variables appearing in the CIT tree.

⁹The correlation of urbanisation with peak court is 0.24 (biserial correlation); with GDP 0.27; and with trademark registrations 0.23. The biserial correlation with capital city is slightly higher at 0.39 while it is somewhat lower for cargo port at 0.12.

6 Rationalization: Institutions, Litigants and Economic Incentives

This Section sets out to rationalize the patterns unearthed in the previous two sections. Before discussing the effect of peak court, GDP, EU trademark registrations, capital and cargo port on the preliminary ruling, we first consider the degree to which the institutional structure of the EU judiciary can account for the magnitude of the regional differential in referral rates. With future research in mind we conclude this Section with a synthesis in the form of a regression model constructed on the basis of the rationalized results of our inductive inquiry.

6.1 Regional Variations and the Basic Institutional Structure of the EU Judiciary

Both our cartography and our CIT model indicate major spatial disparities in referral rates. A small number of regions account for the bulk of preliminary references while most regions see little or no referral activity.

The magnitude of these regional variations can be explained by the institutional features of the mechanism put in place by Article 267 of the Treaty on the Functioning of the European Union, whose wording goes back to the Rome Treaty. Here it is worth quoting the provision in full:

The Court of Justice of the European Union shall have jurisdiction to give preliminary rulings concerning:

- (a) the interpretation of the Treaties;
- (b) the validity and interpretation of acts of the institutions, bodies, offices or agencies of the Union;

Where such a question is raised before any court or tribunal of a Member State, that court or tribunal may, if it considers that a decision on the question is necessary to enable it to give judgment, request the Court to give a ruling thereon.

Where any such question is raised in a case pending before a court or tribunal of a Member State against whose decisions there is no judicial remedy under national law, that court or tribunal shall bring the matter before the Court.

If such a question is raised in a case pending before a court or tribunal of a Member State with regard to a person in custody, the Court of Justice of the European Union shall act with the minimum of delay.

Not only does the provision leave the decision to submit a reference to the judges rather than to the parties. But the language of the provision (“may”) makes it clear that even when EU law is material to the case, the decision to refer a case is a discretionary one for first instance and intermediate courts. Only peak courts are under a duty (“shall”) to request a preliminary ruling from the Court of Justice in such circumstances. Yet the Treaties do not provide for any mechanism to reverse the decision of a peak court failing to uphold its obligation to ask for a preliminary ruling.

Whereas under an appeal mechanism the decision to file an appeal lies with litigants, under a referral procedure the decision to submit a reference belongs to judges. A referral mechanism, therefore, is not well designed to deal with agency issues arising from heterogeneity in the beliefs and preferences of judges across layers of the judicial system. Under an appeal system, losing litigants have the strongest incentive to appeal precisely in those cases in which they expect the appellate court to hold a diverging position (Cameron and Kornhauser, 2006).¹⁰ In contrast,

¹⁰At the same time, the resulting litigant selection effect gives lower courts an incentive to anticipate reversal by conforming to the expected position of the appellate court (Cameron and Kornhauser, 2006, see).

expected divergence works as a disincentive to refer cases under a referral regime. In that sense, it may be said that the nonhierarchical court system comprising the ECJ and national judiciaries reflects a pure team model ([Kornhauser, 1994](#); [Kastellec, 2016](#)), although, as we contend below, informal norms that have emerged from the practices of domestic courts may have put greater pressure on apex tribunals to abide their formal duty under Article 267. What this implies for Article 267 use is that spatial disparities are likely to be more pronounced under a referral regime than under an appellate regime because preference heterogeneity is more likely to come in the way of inter-court cooperation under the former. So, inasmuch as preference heterogeneity may reasonably be assumed to be significant across EU and domestic courts, a good deal of the observed spatial variability may be viewed as resulting from the institutional configuration of the referral regime. These variations are then exacerbated by the clustering of particular courts and litigants (notably those who stand to benefit most from trade liberalization) in a subset of regions, as we now proceed to argue.

6.2 Explaining Peak Court Effect

That the presence of a peak court constitutes, as least for the period 2000-2014, the strongest predictor of referral activity calls for an explanation. Not least since lower courts have been characterised as the “motors of legal integration” ([Alter, 1996](#)). Legal scholars might be tempted to rationalize this finding by pointing to the letter of Article 267 and the formal duty it imposes on courts of last resort. However, we suggest an account that emphasises judicial incentives and resources as well as institutional constraints, which, we believe, makes better sense of the patterns actually observed.

Unlike the EU judiciary, national judiciaries are designed in such a way that they can deal with many of the agency problems arising from preference heterogeneity. Yet here too the team

perspective can account for many of the observed patterns. Judicial systems across EU member states are organized around a three-tier structure, with first instance courts at the bottom, intermediate appeal courts in the middle and supreme courts at the top. To these layers of courts correspond different levels of resources and specialization. First instance courts concentrate on individual dispute resolution and fact-finding, while apex courts focus more on law-finding¹¹ and policy issues that transcend the dispute at hand (Kornhauser, 1994; Kastellec, 2016; Cameron and Kornhauser, 2006). Resources, likewise, differ across the layers of the domestic judicial hierarchy, with top courts typically enjoying greater resources to handle a comparatively smaller caseload (Kornhauser, 1994).¹² Because of their focus on law-finding, peak courts have, *ceteris paribus*, more to win from cooperating with the ECJ—after all preliminary rulings only address points of law. Similarly, greater resources entail that the opportunity cost of writing and submitting a reference are lower. Judges face trade-offs in deciding how to allocate their time (Epstein et al., 2013; Ash and MacLeod, 2015), and time and effort expended in preparing, drafting and submitting a reference are as much time and effort that cannot be used for something else, such as disposing of cases already on the court’s docket or time at home with family and friends. The trade-off, however, will usually be less inhibiting in the case of peak court judges as they have more time and, often, staff support to devote to a single case. The fact is there are thousands of first instance courts across the EU territory but the 70 peak courts that sit at the apex of the 28 national judiciaries now generate more referral activity.

While the team model can explain why peak courts refer more cases at least since the end of the 1990s, it cannot directly explain why it was not so before or at least why the referral activity of lower courts grew faster and their overall share of preliminary references was larger than that

¹¹In most jurisdictions, appeal at the last level is restricted to points of law.

¹²Resources can take the form of law clerks, assistants, access to library and other facilities or simply more judges per work unit.

of both peak and intermediate courts up to that date. The reason for this may be found in the institutional evolution of the preliminary ruling mechanism. Back in the early 1960s, the ECJ was a weak and somewhat obscure judicial entity, as was the preliminary ruling procedure. The first references, therefore, may have had an “experimental” character—judges trying an unconventional solution in an unconventional case. Yet because lower courts are numerous and hear millions of cases every year, there were, despite the low probability, enough such experiments to trigger an institutionalization process (Kingston and Caballero, 2009; Greif and Kingston, 2011; Brousseau et al., 2011; Brousseau and Raynaud, 2011). In keeping with this hypothesis, the greater volatility of first instance court referral rates suggests greater stochasticity in their behaviour. Meanwhile, despite the plain language of Article 267, peak courts had few reasons to take EU law seriously and could ignore their formal obligation to submit references. As cooperation with the ECJ gained more adherents, though, the pressure on peak courts to uphold their duty increased, creating a shared expectation of cooperation.¹³ On this account, the evolution of judicial practices is an illustration of how formal arrangements can interact with informal processes of institutional change. Formally, as laid down in Article 267 of the Treaty on the Functioning of the European Union (TFEU), the preliminary ruling procedure has remained unchanged over the entire lifetime of the European project. Yet the judicial activity that has built around Article 267 has experienced a profound shift.¹⁴

¹³The literature on institutional change has contrasted understandings of institutions as formal rules and as shared beliefs while seeking to clarify the interplay between formal arrangements and informal processes of institutional evolution (Kingston and Caballero, 2009; Greif and Kingston, 2011; Brousseau et al., 2011; Brousseau and Raynaud, 2011). Building on this work, Brousseau and Raynaud (2011) provide a theory of institutional evolution in which promoters of local informal orders compete for adherents. Initially, local orders only apply to a subset of society but can be progressively turned into mandatory, generic orders if they attract enough adherents. This idea makes good sense of the data but is also consonant with the sociological and historical literatures emphasising the role of networks of militant lawyers and law professors in promoting European legal integration when the ECJ was still a relatively obscure judicial body (Cohen and Vauchez, 2007; Rasmussen, 2012).

¹⁴Since the theory posits that peak courts have taken EU law more seriously because of the practice of lower courts, we might want to compare the evolution of the referral activity of peak courts with that of lower courts located in the same region. Aside from sharing the same regional environment, judges in the same region are also more likely to come into contact or to observe each other’s behaviour. So if peak courts followed the lead of lower courts, lower court referral frequency should predict future peak court referral activity. Juxtaposing the two time

While judicial structures matter and go some way to explain high clustering of Article 267 use in regions with peak courts, other indicators with strong predictive power point to the effect of litigation dynamics.

6.3 Explaining the Effect of GDP, Port, Capital and EU Trademark Registrations

Our data has limitations in that we directly observe neither the intermediate and first instance courts that have not referred cases to the ECJ nor the behaviour of individual litigants. Still, we know there are first and intermediate courts in all regions even if these judicial bodies have never been involved in Article 267 proceedings. Thus we can reasonably assume that the predictive power of port, GDP, capital and EU trademark registrations reflects variations in litigation behaviour rather than systematic differences in the preferences and institutional constraints of judicial agents.

Already suggested in previous work on the EU judiciary,¹⁵ the high referral rates associated with the presence of a cargo port is consonant with research linking the use of EU law to trade (Sweet and Brunell, 1998; Pitarakis and Tridimas, 2003; Dyevre and Lampach, 2018) and economic studies that regions with better access to foreign markets benefit more from trade liberalization (Henderson, 2000; Brühlhart, 2011). The removal of barriers to economic exchange and the creation of single market have been central to European integration and remain the focus of EU rules and regulations. As trade hubs, ports generate activity that these rules aim to facilitate. Consequently, when disputes arise, litigants representing the interests of those engaged

series, Figure 17 in the Supplementary Materials shows, as Figure 6, that peak courts initially referred less but eventually overtook the other courts. A Granger causality test with lag $t - 2$ indicates that lower court referral rates predict later peak court referral rates up until the year 2000 ($p = 0.03$).

¹⁵Noting the high number of references coming from the Italian port-city of Genova, Kelemen and Pavone (2016) discussed a possible linkage between port activity and referral rates.

in cross-border trade are likely to invoke EU law to challenge trade barriers. This imply that cases coming from port regions are more likely to focus on EU law provisions aiming at removing obstacles to trade. To validate this interpretation, we compared the texts of cases originated in port regions to those originated in other parts of the EU. We scrapped the texts of all preliminary rulings from the EUR-Lex website and, after applying standard pre-preprocessing steps, we used the following formula to measure how words in each set of cases deviate from expected occurrence:

$$p_{ij} - \frac{p_{ij} + p_{ik}}{2}$$

where p_{ij} is the rate at which word i occurs in the judgments in set j and p_{ik} the rate which it occurs in the judgments in set k . The word cloud in Figure 11 illustrates the words with the largest deviation for the two case sets. Compared to cases from other EU regions, words like “customs”, “product”, “good”, “trade”, “export”, “import” and “animal” feature more frequently in cases coming from port regions. As additional measure of corpus dissimilarity, we took two equal-sized random samples of judgments from the two sets of cases and conducted χ^2 and Mann-Whitney tests. Both tests are strongly significant. Table 1 reports the 20 lexemes with the strongest deviation from the null hypothesis. These largely overlap with the words appearing in Figure 11.

Table 1: Most dissimilar words in random samples of preliminary rulings in cases originating from regions with large cargo ports and regions without large cargo ports.

	Port	No Port	diff
custom	5,838	2,623	3,215
netherland	3,264	618	2,646
product	7,968	5,514	2,454
communiti	7,961	5,544	2,417
unit	3,317	1,001	2,316
mark	4,167	1,872	2,295
good	6,242	3,948	2,294
trade	4,397	2,311	2,086
kingdom	2,696	715	1,981
resid	5,649	3,793	1,856
german	556	1,973	1,417
employ	4,240	2,922	1,318
export	2,576	1,376	1,200
import	3,388	2,239	1,149
anim	1,758	623	1,135
compani	4,684	3,549	1,135
protect	4,975	3,844	1,131
case	14,438	13,395	1,043
sixth	2,791	1,751	1,040
been	6,738	5,746	992
licenc	748	1,736	988

Further evidence that trade drives EU law litigation in port regions is furnished by Figure 12, which compares the share of cases pertaining to customs union provisions between portuary and non-portuary regions using annotations from the CURIA database.

Figure 12: Share of cases pertaining to EU customs union provisions in port regions and other EU regions, 2000-2014. Metric is based on annotations from CURIA database.

The ability of EU trademark registrations to predict referral rates, too, can be linked to trade. Unlike national trademarks, EU trademarks provide protection in all member states. Yet, compared to national trademarks, the application process is more expensive (albeit less than for separate applications in all member states) and more cumbersome (an objection in any member state can defeat the entire application). These characteristics make EU trademarks mostly attractive to businesses that operate in several member states if not EU-wide. Economic work has linked EU trademark registrations to export patterns ([Luini and Mangani, 2004](#)) as well as to innovation ([Mendonça et al., 2004](#); [Flikkema et al., 2010](#); [Gotsch and Hipp, 2012](#)). This suggests that EU trademark users fit the description of the typical winner of trade liberalization.

A regression of referral rates on EU trademark registrations with year and region fixed effects yields a positive and strongly significant effect (Table 2). Given the diversity of goods and services

that may potentially be the object of an EU trademark application, it is hard to determine what issues should be more characteristic of cases originating from regions with high levels of EU trademark registrations. We do see, however, words such as “commercial”, “industrial”, “copyright”, “trade”, “mark” and even “patent” as distinctive of cases coming from these regions (using the cut-off from the CIT model). Cases from these regions also emphasise cross-border judicial cooperation, a topic that falls under the “Area of Security and Justice” in the Treaty on the Functioning of the European Union (Figure 13).

Figure 13: Comparison of most dissimilar words in cases originated from regions with EU trademark registration greater or equal than 1015 and other EU regions. Word size denotes deviation from expected occurrence.

Table 2: Relationship between EU trademark registrations and referral activity. Negative binomial regression models with region and year fixed effects.

	Coefficient	Standard error
Constant	0.9650***	0.1283
Trademark	0.0008***	0.0001
<i>Controlling for year and region fixed effects</i>		
Observation		2752
Log-likelihood		-2534

Note: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$

Although a capital city effect may be thought to reflect the presence of central political institutions and the sort of litigation they give rise to, here too business activity may be a better explanation. In addition to linking European integration to increased geographical clustering (Brühlhart and Koenig, 2006), economic studies have found greater levels of concentration of service sector activity in capital cities (Melchior, 2012). Enshrined in EU treaties is the freedom of establishment which guarantees that services can be marketed across national borders. So capital city effects may reflect the greater concentration of service sector industries in capitals. If this is true, one would expect a greater proportion of cases relating to service sector activities in general. Consistent with this explanation, Figure 14 shows that words like “services”, “freedom”, “establishment”, “telecommunications” and “copyright” occur at a higher rate in cases referred by courts located in capital cities. Note that since the territorial jurisdiction of peak courts extends to the entire country references originating in these courts are excluded from the analysis.

pean Ombudsman are willing to sanction national courts for inappropriate behavior.

As a result, actors who lack the resources to appeal and forum-shop will generate only sporadic litigation. (Conant, 2002, 80)

Because peak courts have jurisdiction over the entire country rather than over the regional unit, it could be that the predictive power of regional GDP is in fact capturing wealth differentials among member states. Yet a non-trivial share of the referrals coming from regions with peak courts are not coming from peak courts but from other courts (as shown in Figure 5 for regions hosting the capital). Table 3 presents binomial regression models with region and year fixed effects that help us untangle several possible rationalizations. The first model evaluates the effect of regional GDP on referral rates by courts other than apex courts. The nonsignificant regression coefficient suggests that the predictive power cannot be explained alone by wealthier litigants bringing more cases before lower courts in the region. The second model relates peak court referrals to national GDP. It reveals a significant but negative association between GDP and peak court referral activity. This result, however, appears to be an artefact of the wealth differential, which in some countries is very large, between regions with peak courts (which tend to be located in richer regions) and the rest of the countries. When we consider all courts but regress referral activity on regional rather than national GDP, we find a positive and significant relationship between wealth and referral rates—which appears to be the relationship captured by the CIT model. Moreover, as the last model shows, this relationship is better captured by a non-linear function—the impact of an additional point of GDP is stronger at lower levels than it is at higher levels of regional wealth.

Table 3: Relationship between economic prosperity and referral activity in regions with peak courts. Table reports coefficients of negative binomial regression models with region and year fixed effects.

	<i>Other Courts</i>	<i>Peak Courts</i>	<i>All</i>	<i>All</i>	<i>All</i>
	<i>Regional GDP</i>	<i>National GDP</i>	<i>National GDP</i>	<i>Regional GDP</i>	<i>Log Regional GDP</i>
Constant	2.9177***	1.7580***	2.0471***	1.8516***	-10.7217***
	(0.5502)	(0.3074)	(0.2726)	(0.3312)	(2.8474)
GDP	-0.00002	-0.00094**	-0.00057**	0.00002*	1.15514***
	(0.00001)	(0.00037)	(0.00027)	(0.00002)	(0.26466)
<i>Controlling for year and region fixed effects</i>					
Observation	355	445	445	355	355
Log-likelihood	-423.60	-713.34	-847.63	-667.43	-659.95

Note: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$

As with harbours, EU trademark registrations and service actor activity in capital cities, this suggests that much of EU law litigation is driven by the haves who stand to win most from trade liberalization.

6.4 Synthesis

With future research in mind, we specify a parsimonious regression model synthesising the insights gained from the foregoing analysis, which we evaluate using cross-validation rather than within-sample statistical significance. As it does not partition the data into disjoint sets, a regression model cannot exactly emulate the process of recursive partitioning. Still, some insights from CIT can be incorporated into the specification of a multi-level regression model. The CIT model depicted in Figure 10 suggests a parsimonious model with five covariates within two interaction

terms:

$$\begin{aligned}
Y_{ijt} = & \exp(\beta_{0jt} + \beta_{1jt} \text{GDP} * \text{Peak Court} \\
& + \beta_{2jt} |1 - \text{Peak Court}| \text{Trademark} * \text{Capital} * \text{Port} \\
& + e_{ijt} + u_{0jt} + r_{00t})
\end{aligned} \tag{3}$$

where $|1 - \text{Peak Court}|$ ensures that the value of each interaction term is zero when the other is positive and vice-versa so that the their coefficient are effectively estimated independently. Against cross-validation, this model yields highly significant coefficients with median RMSE = 2.63. Although it does not perform as well as the CIT and RF models, it is a substantive improvement on the baseline regression model (Figure 15). In addition to more accurate predictions, it displays smaller variance.

Figure 15: Comparison between baseline model, CIT, RF and alternative regression model inspired by CIT.

(a) Alternative regression model for Italian NUTS regions (2000-2014) with and without litigation as additional covariate (N=315).

(b) Alternative regression model for Italian, French and German NUTS regions (2000-2014) with and without Eurofirms as additional covariate (N=713).

Figure 16: Using alternative baseline regression model to assess predictors in smaller data sets compiled by [Kelemen and Pavone \(2016\)](#) and [Kelemen and Pavone \(2018\)](#).

This parsimonious regression model provides a baseline to assess the predictive power of regional factors considered by [Kelemen and Pavone \(2016\)](#) and [Kelemen and Pavone \(2018\)](#) but whose inclusion in our machine learning modelling was precluded by the relative dearth of data. 16 compares the performances of this model on the smaller data sets assembled by [Kelemen and Pavone \(2016\)](#) and [Kelemen and Pavone \(2018\)](#) with and without the two covariates these papers found to be significantly associated with referral activity, namely litigation and the number of law firms specializing in EU law. Cross-validation suggests that adding litigation improves predictive accuracy whereas adding Eurofirms does not. Plausibly, although four of the five variables in the alternative regression model relate to litigation, they do not completely capture its impact. Eurofirms, meanwhile, tend to be located in capital cities or in places with high EU trademark registrations or high GDP, implying that the presence of Euro-law firms is endogenous to these variables.

7 Conclusion

We explored six decades of formal judicial cooperation between EU and national courts. Instead of directly testing hypotheses derived from theory, we used a combination of cartography and machine learning to investigate patterns of referral activity in a more inductive fashion. We uncovered large spatial disparities across EU regions. We found that participation in the referral procedure introduced by the Rome Treaty is highly clustered in politically, economically and legally salient regions. Finally, we undertook to rationalize these results using insights from economic geography and the team theory of adjudication together with additional statistical tests and automated content analysis.

The reach of the EU judiciary is limited by its decentralized, pure team structure. Referral regimes are not designed to deal with agency costs and actors lacking incentives to collaborate with the ECJ may simply shun EU law, leading to considerable spatial disparities in EU law use. Resource constraints and the division of labour within national judiciaries can further explain the growing involvement divide between higher and lower courts. On the litigation side, we suggest that the divide between haves and have-nots, and between the winners and losers of trade liberalization can explain variations between poor and wealthy regions, between regions with large cargo ports and regions with poorer access to international markets and between regions with high and low levels of EU trademark registrations.

The broad narrative emerging from our study thus links the structure of transnational legal institutions to economic forces and global capitalism as the main factors modulating the effectiveness of supranational law. Yet gaps in data availability, whether over courts or litigation, mean there is plenty of room for future research. Court- or even judge-level data would allow for a more fine-grained analysis of variations in EU law use, as would better litigation data.

However, our study, by identifying factors that have strong, out-of-sample predictive power, can serve as guide for future investigation.

References

- Achen, C. H. (2002). Toward a new political methodology: Microfoundations and art. *Annual Review of Political Science* 5(1), 423–450.
- Alter, K. J. (1996). The European Court’s Political Power. *West European Politics* 19(3), 458–487.
- Alter, K. J. (2001). *Establishing the supremacy of European law: The making of an international rule of law in Europe*. Oxford University Press Oxford.
- Alter, K. J. (2012). The global spread of european style international courts. *West European Politics* 35(1), 135–154.
- Alter, K. J. and L. R. Helfer (2010). Nature or Nurture? Judicial Lawmaking in the European Court of Justice and the Andean Tribunal of Justice. *International Organization* 64(04), 563–592.
- Alter, K. J. and J. Vargas (2000). Explaining variation in the use of european litigation strategies: European community law and british gender equality policy. *Comparative Political Studies* 33(4), 452–482.
- Ash, E. and W. B. MacLeod (2015). Intrinsic motivation in public service: Theory and evidence from state supreme courts. *The Journal of Law and Economics* 58(4), 863–913.
- Beim, D., A. V. Hirsch, and J. P. Kastellec (2014). Whistleblowing and Compliance in the Judicial Hierarchy. *American Journal of Political Science* 58(4), 904–918.
- Breiman, L. (2001). Random forests. *Machine learning* 45(1), 5–32.

- Breiman, L. et al. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical science* 16(3), 199–231.
- Brousseau, E., P. Garrouste, and E. Raynaud (2011). Institutional changes: Alternative theories and consequences for institutional design. *Journal of Economic Behavior & Organization* 79(1-2), 3–19.
- Brousseau, E. and E. Raynaud (2011). “climbing the hierarchical ladders of rules”: A life-cycle theory of institutional evolution. *Journal of Economic Behavior & Organization* 79(1-2), 65–79.
- Brühlhart, M. (2001). Evolving geographical concentration of european manufacturing industries. *Review of World Economics* 137(2), 215–243.
- Brühlhart, M. (2011). The spatial effects of trade openness: a survey. *Review of World Economics* 147(1), 59–83.
- Brühlhart, M. and P. Koenig (2006). New economic geography meets comecon. *Economics of Transition* 14(2), 245–267.
- Cameron, C., J. Segal, and D. Songer (2000). Strategic Auditing in a Political Hierarchy: An Informational Model of the Supreme Court’s Certiorari Decisions. *Am. Pol. Sci. Rev.*, 101–116.
- Cameron, C. M. and L. A. Kornhauser (2006). Appeals mechanisms, litigant selection, and the structure of judicial hierarchies. *Institutional Games and the US Supreme Court*, 173–204.
- Carrubba, C. and L. Murrah (2005). Legal integration and use of the preliminary ruling process in the european union. *International Organization* 59, 399–418.

- Chalmers, D. (2000). The positioning of eu judicial politics within the united kingdom. *West European Politics* 23(4), 169–210.
- Cichowski, R. (2007). *The European Court and Civil Society: Litigation, Mobilization and Governance*. Cambridge University Press.
- Cohen, A. and A. Vauchez (2007). Introduction: law, lawyers, and transnational politics in the production of europe. *Law & Social Inquiry* 32(1), 75–82.
- Conant, L. J. (2002). *Justice Contained: Law and Politics in the European Union*. Cornell University Press.
- Cranmer, S. J. and B. A. Desmarais (2017). What can we learn from predictive modeling? *Political Analysis*, 1–22.
- Davies, B. (2012). *Resisting the European Court of Justice*. Cambridge: Cambridge University Press.
- Davies, B. and M. Rasmussen (2012, August). Towards a New History of European Law. *Contemporary European History* 21(Special Issue 03), 305–318.
- Dormann, C. F., J. Elith, S. Bacher, C. Buchmann, G. Carl, G. Carré, J. R. G. Marquéz, B. Gruber, B. Lafourcade, P. J. Leitão, et al. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36(1), 27–46.
- Dyevre, A. and N. Lampach (2017, January). The choice for europe: Judicial behaviour and legal integration in the european union. *Forthcoming* (1).
- Dyevre, A. and N. Lampach (2018). The origins of regional integration: Untangling the effect of trade on judicial cooperation. *International Review of Law and Economics*.

- Eisenberg, T., S. Kalantry, and N. Robinson (2012). Litigation as a measure of well-being. *DePaul L. Rev.* 62, 247.
- Elkins, Z., T. Ginsburg, and J. Melton (2014). Comparative constitutions project. characteristics of national comparative. version 2.
- Epstein, L., W. M. Landes, and R. A. Posner (2013). *The Behavior of Federal Judges: A Theoretical and Empirical Study of Rational Choice*. Harvard University Press.
- Eurobarometer (2016). Eurobarometer data service: Standard and special topic eurobarometer. Gesis Leibzig Institute for Social Sciences.
- Fligstein, N. and A. Stone Sweet (2002). Constructing politics and markets: An institutionalist account of european integration. *American Journal of Sociology* 107(5), 1206–1243.
- Flikkema, M. J., A.-P. de Man, M. J. J. Wolters, et al. (2010). New trademark registration as an indicator of innovation: results of an explorative study of benelux trademark data.
- Fokkema, M., N. Smits, A. Zeileis, T. Hothorn, and H. Kelderman (2017). Detecting treatment-subgroup interactions in clustered data with generalized linear mixed-effects model trees. *Behavior research methods*, 1–19.
- Franzese Jr, R. J. and J. C. Hays (2008). Interdependence in comparative politics: Substance, theory, empirics, substance. *Comparative Political Studies* 41(4-5), 742–780.
- Galanter, M. (1974). Why the "haves" come out ahead: Speculations on the limits of legal change. *Law & society review* 9(1), 95–160.
- Garrett, G. (1995). The politics of legal integration in the European Union. *International Organization* 49, 171–171.

- Gotsch, M. and C. Hipp (2012). Measurement of innovation activities in the knowledge-intensive services industry: A trademark approach. *The Service Industries Journal* 32(13), 2167–2184.
- Greif, A. and C. Kingston (2011). Institutions: rules or equilibria? In *Political economy of institutions, democracy and voting*, pp. 13–43. Springer.
- Hayes, T., S. Usami, R. Jacobucci, and J. J. McArdle (2015). Using classification and regression trees (cart) and random forests to analyze attrition: Results from two simulations. *Psychology and aging* 30(4), 911.
- Heidema, A. G., J. M. Boer, N. Nagelkerke, E. C. Mariman, E. J. Feskens, et al. (2006). The challenge for genetic epidemiologists: how to analyze large numbers of snps in relation to complex diseases. *BMC genetics* 7(1), 23.
- Henderson, J. V. (2000). The effects of urban concentration on economic growth. Technical report, National bureau of economic research.
- Hendry, D. F. and H.-M. Krolzig (2004). We ran one regression. *Oxford bulletin of Economics and Statistics* 66(5), 799–810.
- Hersh, E. D. and C. Nall (2016). The primacy of race in the geography of income-based voting: New evidence from public voting records. *American Journal of Political Science* 60(2), 289–303.
- Hornuf, L. and S. Voigt (2015). Analyzing preliminary references as the powerbase of the european court of justice. *European Journal of Law and Economics* 39(2), 287–311.
- Hothorn, T., K. Hornik, and A. Zeileis (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical statistics* 15(3), 651–674.

- Howe, P. D., M. Mildenberger, J. R. Marlon, and A. Leiserowitz (2015). Geographic variation in opinions on climate change at state and local scales in the usa. *Nature Climate Change* 5(6), 596–603.
- Jacobi, T. (2009a, January). The Role of Politics and Economics in Explaining Variation in Litigation Rates in the U.S. States. *The Journal of Legal Studies* 38(1), 205–233.
- Jacobi, T. (2009b). The role of politics and economics in explaining variation in litigation rates in the us states. *The Journal of Legal Studies* 38(1), 205–233.
- Kastellec, J. (2016). The judicial hierarchy: A review essay.
- Keefer, P. (2012). Database of political institutions: Changes and variable definitions. The World Bank: Development Research Group.
- Kelemen, D. and T. Pavone (2016). Mapping european law. *Journal of European Public Policy* 23(8), 1118–1138.
- Kelemen, D. and T. Pavone (2018). The political geography of european legal integration: Visualizing institutional change in the european union. *World Politics* 70(3).
- Kelemen, R. D. (2011). *Eurolegalism: the transformation of law and regulation in the European Union*. Harvard University Press.
- Kelemen, R. D. (2012). Eurolegalism and democracy. *JCMS: Journal of Common Market Studies* 50(s1), 55–71.
- Kingston, C. and G. Caballero (2009). Comparing theories of institutional change. *Journal of Institutional Economics* 5(2), 151–180.

- Kornhauser, L. A. (1994). Adjudication by a resource-constrained team: Hierarchy and precedent in a judicial system. *S. Cal. L. Rev.* 68, 1605.
- Lessmann, C. (2014). Spatial inequality and development—is there an inverted-u relationship? *Journal of Development Economics* 106, 35–51.
- Luini, L. and A. Mangani (2004). Trademarks, product variety, and economic activity in Italy and Europe.
- Melchior, A. (2012). East–west integration: A geographical economics approach. In *EU Eastern Neighborhood*, pp. 23–44. Springer.
- Mendonça, S., T. S. Pereira, and M. M. Godinho (2004). Trademarks as an indicator of innovation and industrial change. *Research Policy* 33(9), 1385–1404.
- North, D. C. (1990). *Institutions, institutional change and economic performance*. Cambridge university press.
- Pinker, S. (2011). *The better angels of our nature: The decline of violence in history and its causes*. Penguin UK.
- Pitarakis, J. and G. Tridimas (2003). Joint dynamics of legal and economic integration in the European Union. *European Journal of Law and Economics* 16, 357–368.
- Rasmussen, M. (2012). Establishing a Constitutional Practice of European Law: The History of the Legal Service of the European Executive, 1952–65. *Contemporary European History* 21(Special Issue 03), 375–397.
- Rodríguez-Pose, A. and R. Ezcurra (2009). Does decentralization matter for regional disparities? a cross-country analysis. *Journal of Economic Geography* 10(5), 619–644.

- Rodríguez-Pose, A. and N. Gill (2006). How does trade affect regional disparities? *World Development* 34(7), 1201–1222.
- Romeu, F. R. (2006). Law and politics in the application of ec law: Spanish courts and the ecj 1986-2000. *Common Market L. Rev.* 43, 395.
- Sala-i Martin, X. (1997). I just ran two million regressions. *American Economic Review* 87(2), 178–183.
- Schneider, G. and L.-E. Cederman (1994). The change of tide in political cooperation: A limited information model of european integration. *International Organization* 48(4), 633–662.
- Songer, D. R., S. B. Haire, and R. S. Sheehan (1999). Do the "haves" come out ahead over time? applying galanter's framework to decisions of the us courts of appeals, 1925-1988. *Law & Society Review* 33(4), 811.
- Sweet, A. S. and T. L. Brunell (1998). Constructing a supranational constitution: Dispute resolution and governance in the european community. *American Political Science Review* 92(1), 63–81.
- Sweet, T. A. S. and T. L. Brunell (1999). Data set on preliminary references in ec la. Robert Schuman Centre, European University Institute.
- Varian, H. R. (2014). Big data: New tricks for econometrics. *The Journal of Economic Perspectives* 28(2), 3–27.
- Vachez, A. (2010). The transnational politics of judicialization. van gend en loos and the making of eu polity. *European Law Journal* 16(1), 1–28.

Vink, M., M. Claes, and C. Arnold (2009). Explaining the use of preliminary references by domestic courts in eu member states: A mixed-method comparative analysis. Conference paper Judicial Politics in the EU and Beyond 11th Biennial Conference of the European Union Studies Association.

Weiler, J. H. H. (1991). The Transformation of Europe. *Yale L.J.*, 2403–2483.

Wind, M., D. S. Martinsen, and G. P. Rotger (2009). The uneven legal push for europe questioning variation when national courts go to europe. *European Union Politics* 10(1), 63–88.

Zweynert, J. (2009). Interests versus culture in the theory of institutional change? *Journal of Institutional Economics* 5(3), 339–360.

Supplementary Materials

Table 4: List of variables

Variable name	Explanation	Type of data	Unit	Source
Referral _{<i>i,j,t</i>} (Dependent variable)	number of referrals submitted by referent court located in NUTS2 level <i>i</i> of country <i>j</i> at time <i>t</i>	discrete		Sweet and Brunell (1999)
<i>Regional level data</i>				
GDP _{<i>i,j,t</i>}	gross domestic product at current market prices by NUTS2 level <i>i</i> and country <i>j</i> at time <i>t</i>	continuous	million Euro	Eurostat (2017a)
Employment _{<i>i,j,t</i>}	number of employed persons by NUTS2 level <i>i</i> and country <i>j</i> at time <i>t</i>	continuous	Thousand	Eurostat (2017a)
Unemployment _{<i>i,j,t</i>}	number of unemployed persons by NUTS2 level <i>i</i> and country <i>j</i> at time <i>t</i>	continuous	Thousand	Eurostat (2017a)
Tourists _{<i>i,j,t</i>}	arrivals at tourist accommodation establishments by NUTS2 level <i>i</i> and country <i>j</i> at time <i>t</i>	continuous	number	Eurostat (2017a)
Population _{<i>i,j,t</i>}	population on 1 January by NUTS2 level <i>i</i> and country <i>j</i> at time <i>t</i>	continuous	number	Eurostat (2017a)
Vehicles _{<i>i,j,t</i>}	stock of vehicles by NUTS2 level <i>i</i> and country <i>j</i> at time <i>t</i>	continuous	number	Eurostat (2017a)
Victims _{<i>i,j,t</i>}	victims on road accidents by NUTS2 level <i>i</i> and country <i>j</i> at time <i>t</i>	continuous	number	Eurostat (2017a)
Trade mark _{<i>i,j,t</i>}	European Union Trade mark (EUTM) by NUTS2 level <i>i</i> and country <i>j</i> at time <i>t</i>	continuous	number	Eurostat (2017a)
Primary income _{<i>i,j,t</i>}	allocation of primary income account of households by NUTS2 level <i>i</i> and country <i>j</i> at time <i>t</i>	continuous	million euro	Eurostat (2017a)
Life expectancy _{<i>i,j,t</i>}	Life expectancy by NUTS2 level <i>i</i> and country <i>j</i> at time <i>t</i>	continuous	year	Eurostat (2017a)
Urban _{<i>i,j,t</i>}	percentage of households living in urban area by NUTS2 level <i>i</i> and country <i>j</i> at time <i>t</i>	continuous	percentage	Eurostat (2017a)
Migration _{<i>i,j,t</i>}	net migration by NUTS2 level <i>i</i> and country <i>j</i> at time <i>t</i>	continuous	number	Eurostat (2017a)

Table 4: List of variables

Variable name	Explanation	Type of data	Unit	Source
Maritim $e_{i,j,t}$	dummy variable taken the value 1 if the NUTS2 region i in country j at time t has a large cargo port and 0 otherwise	categorical		Eurostat (2017b)
Peak Court i,j	dummy variable taken the value 1 if a top court is in the NUTS2 region i in country j and 0 otherwise	categorical		NA
Capital i,j	dummy variable taken the value 1 if the referent court is in the capital of country j located in the NUTS2 region i and 0 otherwise	categorical		NA
<i>Country level data</i>				
Membership Duration $_{j,t}$	number of years elapsed since accession to the European Union	continuous		NA
Government Debt $_{j,t}$	public debt by country j at time t	continuous	million Euro	Eurostat (2017c)
Carbon $_{j,t}$	carbon emission (CO ₂) by country j at time t	continuous	kg	Worldbank (2017a)
Corruption $_{j,t}$	corruption perceptions index by country j at time t	continuous	index	Transparency International (2017)
Decentralisation $_{j,t}$	degree of fiscal decentralisation between the central government and the sub-national entity level by the use of measures of revenue, internal government transfers and expenditure by country j at time t	continuous	percentage	Eurostat (2017d)
Voice $_{j,t}$	voice and accountability by country j at time t	continuous	score	Worldbank (2017b)
Political Stability $_{j,t}$	political stability and absence of violence by country j at time t	continuous	score	Worldbank (2017b)
Voice $_{j,t}$	voice and accountability by country j at time t	continuous	score	Worldbank (2017b)
Gov. Eff. $_{j,t}$	government effectiveness by country j at time t	continuous	score	Worldbank (2017b)
Regulatory $_{j,t}$	regulatory quality by country j at time t	continuous	score	Worldbank (2017b)
Rule $_{j,t}$	rule of law by country j at time t	continuous	score	Worldbank (2017b)
EU Support $_{j,t}$	net percentage difference between respondents saying that EU membership is a good thing and saying that EU membership is a bad thing	continuous	percentage	Eurobarometer (2016)
Political Fragmentation $_{j,t}$	number and ideological distance among domestic veto players	discrete		Keefer (2012)
Monism $_{j,t}$	dummy variable taking the value one if treaties are granted superiority over domestic legislation to ordinary legislation in the country i and zero otherwise	categorical		Comparative Constitution Project Elkins et al. (2014)
Constitutional Review $_{j,t}$	dummy variable which takes the value one if member state has adopted Kelsenian (i.e. centralized) model of constitutional review and zero otherwise	categorical		Comparative Constitution Project Elkins et al. (2014)

Note: NA means that no data source is necessary to compile the variable

Table 5: Regional breakdown of the data

Country	Number of Administrative Units NUTS2 regions		NUTS Code	Observations
Austria	9	States (Bundesländer)	AT	
Belgium	11	Provinces (Provinces)	BE	Including Brussels
Denmark	5	Regions (Regioner)	DK	
Finland	5	Large Areas (Suuralueet/Storomraden)	FI	
France	22	Regions (Régions)	FR	Excluding Overseas Department
Germany	39	Government Regions (Regierungsbezirke)	DE	
Greece	13	Regions (Perifereies)	EL	NUTS2 regions 2000
Ireland	2	Regions	IR	
Italy	21	Regions (Regioni)	IT	Trentino-Alto Adige/Südtirol is split in two region
Luxembourg	1		LU	
Netherlands	12	Provinces (Provincies)	NL	
Spain	19	Autonomous Communities and Cities (Comunidades y ciudades autónomas)	ES	
Portugal	7	Autonomous Regions	PT	Including Acores and Madeira
Sweden	8	National Areas (Riksomraden)	SE	
United Kingdom	40	Counties	UK	
Cyprus	1		CY	
Estonia	1		EE	
Hungary	7	Statistical Regions (Tervezési-statisztikai régiók)	HU	
Latvia	1		LV	
Lithuania	1		LT	
Malta	1		MT	
Poland	16	Voivodeships (Województwa)	PL	
Czech Republic	8	Oblasts (Oblasti)	CZ	
Slovakia	4	Oblasts (Oblasti)	SK	
Slovenia	2	Macroregions (Kohezijske regije)	SI	
Bulgaria	6	Regions (Rajoni za planirane)	BG	
Romania	8	Regions (Regiuni)	RO	
Croatia	2	Regions	HR	

Table 6

	Port	No Port	diff
custom	5,838	2,623	3,215
netherland	3,264	618	2,646
product	7,968	5,514	2,454
communiti	7,961	5,544	2,417
unit	3,317	1,001	2,316
mark	4,167	1,872	2,295
good	6,242	3,948	2,294
trade	4,397	2,311	2,086
kingdom	2,696	715	1,981
resid	5,649	3,793	1,856
german	556	1,973	1,417
employ	4,240	2,922	1,318
export	2,576	1,376	1,200
import	3,388	2,239	1,149
anim	1,758	623	1,135
compani	4,684	3,549	1,135
protect	4,975	3,844	1,131
case	14,438	13,395	1,043
sixth	2,791	1,751	1,040
been	6,738	5,746	992
licenc	748	1,736	988

Figure 17: Referral activity in peak court regions. The dashed blue line shows the time series for peak courts. The red solid line shows the time series for intermediate and first instance courts.