The Choice for Europe: Judicial Behaviour and Legal Integration in the European Union

Arthur Dyevre*and Nicolas Lampach†

Abstract

Extending a dataset compiled by Sweet and Brunell (1998b), we investigate the factors influencing the choice of domestic judges to pass on cases to the Court of Justice of the European Union. While EU judicial scholars have typically relied on integration theory or ad hoc theories of adjudication to explain the conduct of judicial actors, we present an account of judicial decision making in the context of the preliminary ruling mechanism more consonant with recent theoretical advances in judicial behaviour research. Based on a Bayesian framework, our estimation strategy addresses issues of endogeneity and distributional assumptions neglected in the previous literature. We find that the rate at which domestic courts refer cases is influenced by litigation volume, familiarity with EU law and political fragmentation, but neither by monism, nor by intra-EU trade. The paper also provides the first full-scale analysis of the impact of enlargement on preliminary references.

Keywords: Judicial behaviour, European Court of Justice, Legal Integration, Bayesian statistics

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

Acknowledgment The authors gratefully acknowledge financial support by the project "EUTHORITY" hosted by KU Leuven, funded through European Research Council (ERC grant). We are grateful to Tomas Adamec, Angelina Atanasova, Gilian Bens, Monika Glavina, Frauke Petra Hein, David Ketch and Anna Maria Tonikidou. We also thank the participants at the Expert Seminar on "*The Deliberative Performance of Constitutional Courts: an Empirical Perspective*" for their useful comments.

* *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:

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

1 Introduction

International legal regimes vary significantly in their ability to engage infra-state actors and to shape domestic policy outcomes. Differences in effectiveness seem to hinge in large on the ability of international adjudicators to forge compliance partnerships with national judiciaries (Keohane et al., 2000; Downs et al., 1996; Huneeus, 2013). In what is arguably the world's most deeply integrated international regime, the European Union, close cooperation between EU courts and national courts is widely credited for creating the conditions that have allowed the effective enforcement of EU laws and regulations in the domestic realm (Gabel et al., 2012; Stone Sweet, 2004; Sweet and Brunell, 1998b). An interlocutory procedure introduced by the Rome Treaty, in particular, is believed to have been instrumental in permitting the transformation of the European Union legal order. Established by Article 267[1] of the Treaty, the preliminary ruling procedure provides a formal setting for domestic judges to refer legal questions to the European Court of Justice. In addition to supplying the Court of Justice with a steady stream of cases, it has furnished domestic judges with authoritative legal doctrines to challenge domestic policies (Stone Sweet, 2004; Alter, 2001; Weiler, 1994, 1991). As illustrated in Figure 1, references have risen dramatically since the inception of the integration process, reflecting the rise, expansion and consolidation of the EU legal system.

Yet, while the preliminary reference mechanism has indubitably grown in popularity, there exist significant disparities among national courts as regards the frequency with which they choose to refer legal questions. Plotted in Figure 2 is the distribution of annual reference counts broken down by Member State. In early years of membership, courts typically send few or no references (except for Austrian courts, which apparently hit the ground running when Austria joined the EU in 1995). Only after some years in the EU do judges start to embrace the practice and reference numbers begin to take off. Yet, even after decades of membership, Belgian courts, for example, still refer more cases than their Danish counterparts in a typical year.

[1] Numbering has been altered by repeated Treaty revisions (although these have left the original wording intact). Article 177 became Article 234 following the ratification of the Amsterdam Treaty, which the Lisbon Treaty changed again to the current numbering.

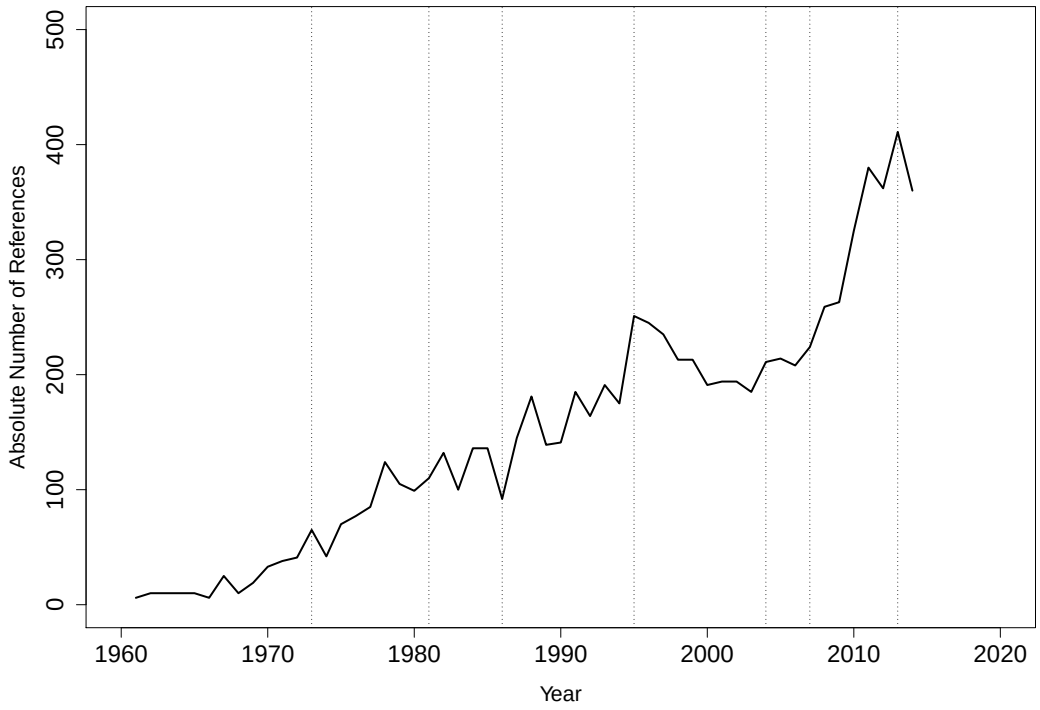


Figure 1: Article 267 References, 1961-2014

Note: Dashed lines denote enlargement years.

Why this is a question that has spawned a fair amount of scholarship. Researchers have investigated a wide array of possible explanations, from patterns of cross-border trade to religious affiliation and legal norms (Sweet and Brunell, 1998b; Carrubba and Murrah, 2005a; Vink et al., 2009; Wind et al., 2009a). These efforts, though, have not converged towards a set of clear conclusions over the determinants of judicial behaviour in this area. Some authors have found intra-EU trade to be positively associated with more frequent references (Sweet and Brunell, 1998b; Carrubba and Murrah, 2005a). But others have reached the opposite conclusion (Wind et al., 2009a; Kelemen and Pavone, 2016). Similarly, some have found monism to be a significant predictor of referral behaviour (Carrubba and Murrah, 2005a; Vink et al., 2009), but other studies have cast doubt on this finding (Hornuf and Voigt, 2015). The same for public support for EU membership, which one analysis found to be a significant predictor of referral behaviour (Carrubba and Murrah, 2005a), only to be rebutted by a later study (Hornuf and Voigt, 2015). To be sure, these differences may be attributed, in part, to differences in modelling assumptions and data coverage. Scholars have crunched their data using an eclectic set of techniques, including cross-sectional OLS regression (Sweet and Brunell, 1998b; Fligstein and Stone Sweet, 2002; Vink

et al., 2009), pooled cross-section (Sweet and Brunell, 1998b; **?**), variants of panel setups (Hornuf and Voigt, 2015; Carrubba and Murrah, 2005a) and even, in one case, Boolean analysis (Vink et al., 2009).

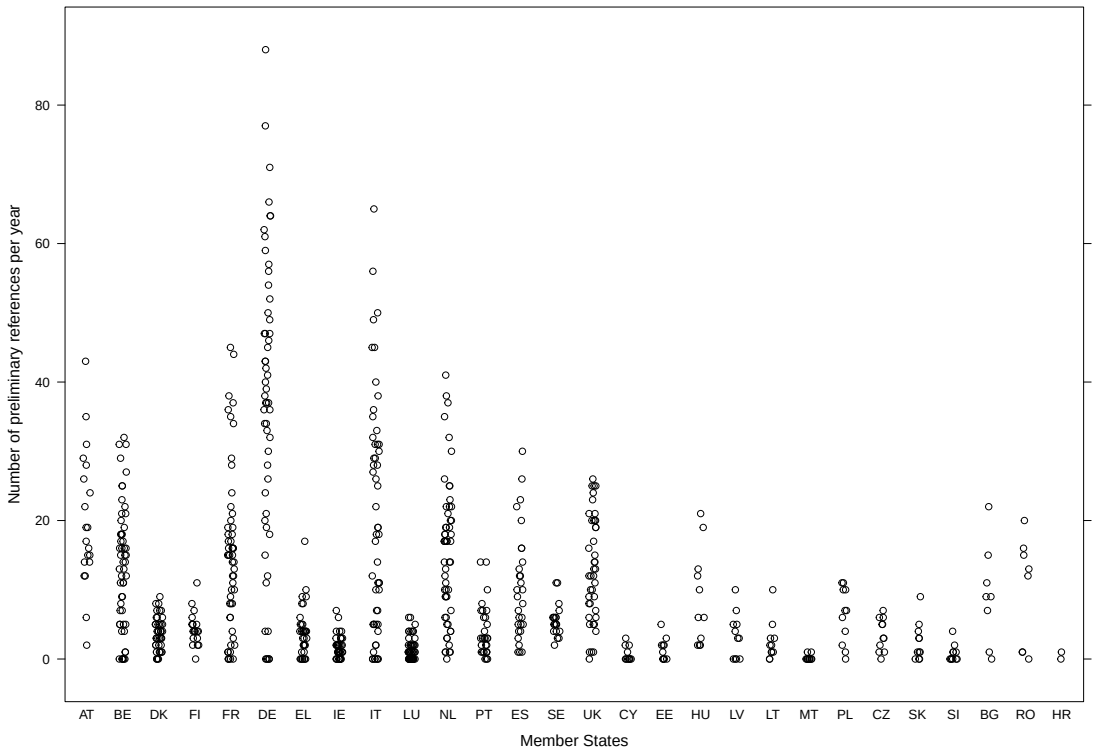


Figure 2: Article 267 References Broken Down by Member State, 1961-2014

Variations in sample size, meanwhile, may explain why studies covering a longer timespan have generally found the relationship between trade and the preliminary ruling system to be less robust, or even insignificant, as shown in Figure 3. The problem, however, runs deeper. And it is not one that can be fixed by simply collecting more data (although it can certainly help). Nor is it only a matter of addressing the methodological flaws (the operationalization of the variable "monism" in Carrubba and Murrah (2005a) and Vink et al. (2009) is a case in point) that make some analyses impossible to replicate. Indeed, afflicting these studies is a more general shortcoming that pertains to their lack of theoretical elaboration and micro-foundations. More concerned with the theme of European integration than with the theory of judicial decision making, the field has failed to evolve an account of judicial motivation sufficiently specific to guide the process of hypothesis derivation and evaluation. As a result, and it should not come as a surprise, no coherent picture of what drives the judges' choice for Europe–i.e. their propensity to submit legal questions to the Court of Justice–has emerged from the past two decades of research.

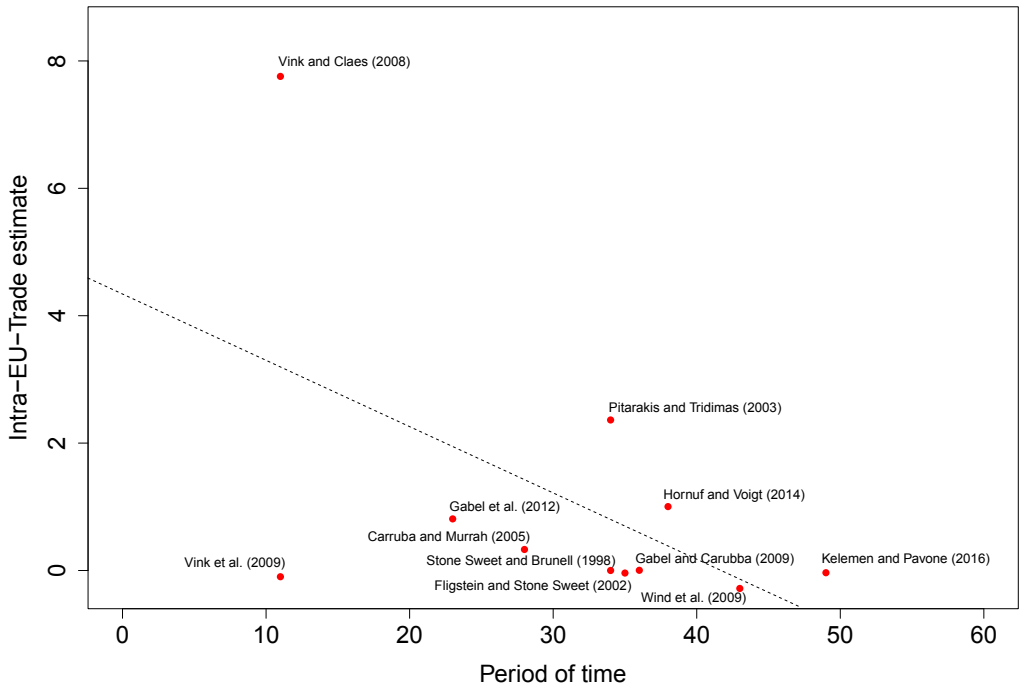


Figure 3: Estimated Effect of Trade on Preliminary References and Time Coverage of Previous Studies

This paper makes three contributions to the preliminary ruling literature. First, we place our analysis within the broader field of judicial behaviour research. Where previous inquiries have generally relied on integration theory and ignored research on judicial institutions conducted in other settings, our analysis seeks to incorporate recent theoretical advances in judicial decision making research while providing more rigorous micro-foundations for the study of adjudication in the preliminary ruling environment. Second, we adopt a Bayesian approach that, we argue, does not only offer more robust estimates but also allows us to address methodological issues which have not received sufficient attention in the literature. We discuss different modelling strategies employed by previous authors and compare alternative possible priors. Our Bayesian approach also addresses issues of endogeneity neglected in previous research via instrumental variable estimation. Finally, extending the dataset originally compiled by (Sweet and Brunell, 1998b) to 2014, our analysis offers the most comprehensive assessment to date of the behaviour of domestic courts towards the preliminary ruling mechanism in the countries that joined the EU during the decade past. In the end, we find evidence that cross-national variations in referral behaviour are affected by litigation volume, familiarity with EU law and political fragmentation, but not by monism, public support for integration or intra-EU trade. We show that the correlation

between trade and references is essentially driven by Germany and disappears when we eliminate endogeneity. We also reject the notion that courts in new member states behave differently vis-à-vis EU law and the preliminary ruling system.

The paper is organized as follows. Section 2 reviews existing literatures on EU judicial politics and adjudication theory and presents our theory and hypotheses. Drawing on recent judicial behaviour research, we specify individual judicial preferences and how these relate to institutional and political constraints. Section 3 describes the data and accompanying covariates. Section 4 then sets out our estimation strategy. Finally, Section 5 presents and discusses the results our analysis. We conclude with some brief thoughts on the direction of future research.

2 Judicial Cooperation and Judicial Behaviour

The field of judicial behaviour research has seen significant theoretical advances over the past ten years. First, judicial scholars have moved beyond the unproductive opposition between the attitudinal and the legal model in favour of richer and more realistic conceptions of judicial decision making (Stephenson, 2009; Epstein and Knight, 2013). Where the theoretical discussion used to revolve almost exclusively around policy goals and legal rules, it is now widely recognised that judicial conduct is modulated by a broader set of considerations, including preference for leisure and social prestige (Epstein et al., 2013; Epstein and Knight, 2013). Second, efforts have been made to model with greater rigour the trade-offs judges face when deciding how to allocate their limited resources (Epstein et al., 2013; Ash and MacLeod, 2015). Third, as formal modellers have continued to explore the strategic constraints variously arising from the number and configuration of legislative veto-players, public attitudes towards the courts and the hierarchical structure of the judiciary (Rogers et al., 2006; Vanberg, 2005; Clark, 2009), attempts have been made to systematically integrate these various strands of strategic analysis within a more general theory of judicial decision making Dyevre (2010); Hammond et al. (2005). Finally, another important evolution in the field has been spurred by the behavioural revolution in decision theory, as scholars, borrowing insights from behavioural psychology, have begun to explore how cognitive biases affect judicial outcomes (Chen et al., 2016; Danziger et al., 2011).

These theoretical developments have been slow to enter the social-scientific literature on the EU legal system. This reflects in part the sway grand theories of integration–neo-functionalism and intergovernmentalism–have long held over the field of EU studies (Stone Sweet and Brunell, 2012; Sweet and Brunell, 1998b; Garrett, 1995; Burley and Mattli, 1993). Yet even those critical of both neo-functionalist and intergovernmentalist narratives of European legal integration have disregarded the insights delivered by research on courts and judges in other jurisdictions, electing, instead, to devise their own, EU-tailored theories of adjudication (Wind et al., 2009a). This

state of affairs has had two pernicious consequences. For one, the artificial divide between the field of EU judicial studies and mainstream judicial behaviour research has worked to impede cross-illumination and cross-fertilization and, thus, greater scientific consilience among inquiries conducted in different contexts.[2] For another, it has forestalled theoretical progress in an area of research that lacks micro-foundations. Neo-functionalist accounts of legal integration link trade flows, litigation and dispute resolution, but the causal mechanisms supposed to connect these elements are underspecified. Stone Sweet and Brunell argue that judges seek an efficient way to dispose of their caseload: "to go home at the end of the day having disposed of more, rather than fewer work-related problems" (Sweet and Brunell, 1998a, 73). Yet it is hard to see how this could possibly motivate domestic judges to refer questions to the Court of Justice. Indeed, the costs associated with the procedure for the referring judge (which include the time and intellectual effort required to write the reference and await the ECJ's preliminary ruling) means sending a reference will usually add to, rather than subtract from her net workload. Common to many papers is the tendency to roll up assorted hypotheses from qualitative and quantitative studies without much attempt to integrate them within a coherent theory of judicial decision making (e.g. Vink et al., 2009; Carrubba and Murrah, 2005a). Finally, there is the extreme case of the paper throwing everything at the data–including the proverbial kitchen sink (which,in that case, takes the form of an indicator for the Roman Catholic share of the population in 1980)–seemingly in the hope that some variables will come out as statistically significant (see Hornuf and Voigt, 2015). Whatever merits these studies may have in other respects, it is clear that they contribute little to the construction of a systematic picture of judicial choice making.

In the remainder of this section, we try to articulate a conception of judicial decision making in the preliminary ruling system more consistent with the picture of judging that has emerged from broader judicial behaviour research. We follow Dyevre (2010) in relating, deriving and distinguishing micro-, meso- and macro-level hypotheses (Table 1).

Table 1: Determinants and Constraints of Judicial Behaviour.

Level of Analysis	Variables
Macro-Level	Public Support, Political Fragmentation
Meso-Level	Caseload, Monism, Judicial Review
Micro-Level	Information, Individual Preferences

[2]For a discussion of consilience as a criterion for theory choice and scientific progress see Thagard (1978).

2.1 Micro-Level Determinants: Legal Knowledge and Judicial Preferences

Starting with micro-level determinants, we assume that domestic judges are motivated by preferences for (1) professional reputation, (2) policy outcomes consistent with their ideological leaning and (3) leisure time. Preference for professional reputation means that, other things being equal, judges will prefer the course of action that enhances or, at least, maintains the perception (and maybe also inner sentiment) that they are doing their job well (Ash and MacLeod, 2015; Epstein et al., 2013). Other things being equal, judges will also prefer outcomes that are closer to their ideological ideal point. With regard to the choice to submit a preliminary reference, we assume that policy preferences may influence judges both on the pro vs anti-integration dimension and on the left-right dimension. In other words, a judge may want to refer a legal question out of a desire to promote deeper integration and greater compliance with EU law. Alternatively though, a judge otherwise indifferent to EU law may view the preliminary ruling mechanism as an opportunity to challenge domestic laws and doctrines she considers objectionable. Finally, preference for leisure time implicates that, other things being equal, judges will prefer the course of action that maximizes the time left for non-judicial activities, including family time, rest, hobbies and non-judicial work such as teaching or writing (e.g. law review articles). Now, to the extent that judges have limited resources in time and staff support, they face important trade-offs in seeking to maximizes these preferences. Raising one's reputation will usually entail more (judicial) work and less leisure time; the same as advancing policy goals departing from the status quo may come at the expense of reputation unless one invests sufficient efforts in crafting persuasive legal arguments, which will also in turn affect leisure time (Stephenson, 2009). Assuming decreasing marginal satisfaction (every additional unit of time devoted to either work or leisure increases the judge's satisfaction but not as much as the previous unit), judicial outcomes will thus reflect the judges' priority given their current time allocation. A small workload will make leisure seem a less pressing need, leading judges to devote more of their time to policy and reputational considerations. A large one, by contrast, will drive up the opportunity cost of engaging in policy-oriented and reputation-enhancing activities and make quick and efficient case disposition the priority.

Conceptualizing judicial decision making as an optimisation problem (Ash and MacLeod, 2015) helps shed light on the costs associated with the decision to submit a preliminary reference. The cost directly associated with a referral has both an external and an internal component. The external component is determined by the Court of Justice. It reflects the time the Court of Justice makes the domestic judge wait before rendering its preliminary ruling[3] together with the (reputational or policy) risk entailed by an adverse ECJ decision.[4] The internal component, on

[3] Average time between reference referral and preliminary ruling has oscillated between 18 and 24 months.

[4] Examples of adverse decisions include decision of dismissal (reputationally costly) or decisions that do not

the other hand, reflects the investment in time and mental effort required to prepare and write the reference and its magnitude depends on the domestic judge's ability and knowledge. In that regard, it appears reasonable to presume that low ability and poor knowledge of EU law will make the internal cost component larger, whereas high ability and good knowledge of EU law will make it smaller. Together, these two cost components also determine the opportunity cost in leisure time or judicial output foregone resulting from the decision to refer.

Were we able to measure them accurately, the judges' policy preferences and workload would probably account for some of the variance observed in the data. Some member states may have a higher proportion of pro-integration judges, while others may have smaller judge/caseload ratio or a higher proportion of lazy judges. Because only a tiny fraction of the million cases litigated every year in the EU give rise to preliminary references, it seems sensible to posit that the opportunity cost of submitting a preliminary reference generally overweight the policy and reputational benefits domestic judges might otherwise anticipate from a preliminary ruling. This probably holds true even if we consider only the external component of the cost associated with a referral. Testing this hypothesis, of course, would require EU-wide, longitudinal measures of judicial attitudes along with reliable indicators of prevailing judge/workload ratios–measures that have yet to be developed.[5] What we can test, however, is the extent to which exposure to EU law reduces the internal cost component and, thereby, the opportunity cost of a referral:

> $Hypothesis_1$: Longer exposure to EU law will result in higher propensity to use the preliminary reference system.[6]

In more explicitly behavioural terms, greater familiarity reduces the cognitive strain that EU law will otherwise elicit, making judges more comfortable to use the preliminary ruling mechanism.[7]

2.2 Meso-Level Incentives and Constraints: Monism, Judicial Review, Litigation and Trade

Aside from workload and knowledge of EU law, a judge's preference-maximizing behaviour is potentially subject to a number of meso-level constraints arising from the configuration of the court system and the substance of legal rules. One way to analyse the constraining effect of legal rules on judging is as effectively imposing an argumentation cost on the judge seeking to deviate from these rules (Stephenson, 2009). Unless the judge comes up with a persuasive argument to

comport with the domestic judge's policy agenda.

[5] Existing workload indicators are difficult to trust and have limited temporal coverage see CEPEJ data.

[6] Previous studies have tested similar formulations of this hypothesis but not on the basis of the theoretical derivation presented here.

[7] Surveys and interviews conducted in Germany, Poland and the Netherlands show that judges often report complexity and lack of knowledge as a reason both to avoid using EU law in resolving disputes and to steer clear of the preliminary ruling mechanism, see Mayoral et al. (2014) and Jaremba (2013). On the role of processing fluency in cognition and decision making see Alter and Oppenheimer (2009).

justify her deviation from an established rule of law, her decision will attract criticism, resulting in a reputational loss. Crafting a persuasive legal argument, however, requires an investment in time and mental effort usually commensurate with the magnitude of the deviation. The size of the argumentation cost thus defines the legal rule's constraining force.

Such a view of the effect of legal rules would appear to sit well with the contention that a monist legal tradition, by explicitly granting supremacy and direct effect to international law, makes it it easier for national judges to appeal to the authority of a supranational adjudicator (Carrubba and Murrah, 2005a). This argument, however, overlooks the fact that even in a dualist system, the argumentation cost to use the preliminary ruling mechanism will fall dramatically once a domestic precedent articulating the rationale for EU law supremacy and direct effect exists. Moreover, while highly contested in the early years of the European project, the doctrines of supremacy and direct effect now constitute an integral part of what is known in EU jargon as the "Community acquis", i.e. the body of principles, rules and regulations to which countries must commit prior to accession. This suggests that domestic courts in new Member States faced little need to persuade their audiences that their legal tradition allowed them to refer legal questions to the Court of Justice. Our expectation, therefore, is that monism will bear no influence on referral behaviour:

> $Hypothesis_2$: Monism will have no influence on the judges' propensity to use the preliminary ruling system.

Another set of hypotheses discussed in the preliminary ruling literature relates to the courts' formal powers to exercise judicial review under domestic law (Hornuf and Voigt, 2015; Carrubba and Murrah, 2005a). In some Member States, the Netherlands for instance, courts do not have the power to set aside statutory legislation on constitutional grounds. In other Member States, such as Portugal, all courts enjoy this right. In most Member States, though, the right exists but is the exclusive preserve of a specialized constitutional court. Scholars, starting with Joseph Weiler, have argued that where courts do not have the power to strike down legislation under domestic law policy-minded judges have a strategic incentive to embrace EU law as as means to acquire this power (Weiler, 1991, 1994). This empowerment theory implies that Member States without judicial review or where judicial review is centralized in a constitutional court will see more courts submitting more frequent references, whereas Member States with decentralized judicial review will see less.

> $Hypothesis_3$: Judges with more extensive powers of judicial review will have a lower propensity to use the preliminary ruling system than judges in Member States with no judicial review or where judicial review is more limited in scope.

A related, yet distinct theoretical proposition stipulates that the existence of a centralized system of judicial review will incentivize policy-minded judges sitting on ordinary courts to counter the influence of the constitutional court by making more frequent use of EU law (Alter, 1998). Constitutional court judges, obviously, have relatively less to gain from legal integration (they already have the power invalidate statutory legislation). Yet the court competition hypothesis predicts that their reluctance to submit references will be more than compensated by the enthusiasm of other domestic courts:

> $Hypothesis_4$: Compared to systems without judicial review or with decentralized judicial review, the existence of centralized judicial review will result in higher propensity to use the preliminary ruling mechanism.

Now, however keen a judge may be to write and submit references, her ability to do so will be to a large degree determined by the odds that a litigant turns up with the right sort of legal case, i.e. a case bearing some relation to EU law. Partly because many early ECJ landmark cases–including *Van Gend en Loos*, *Dassonville* and *Cassis de Dijon*–originated in trade disputes, it made intuitive sense to hypothesise that trade-related cases were more likely to give rise to preliminary references. Neo-functionalism put theoretical flesh on this intuition by specifying how EU law created incentives for businesses involved in cross-border trade to use domestic courts to challenge trade barriers (Sweet and Brunell, 1998b; Stone Sweet, 2004). Yet there are two reasons to doubt the claim that more cross-border trade causes judges to refer more. The first is reverse causality. It is easy to argue that references, by leading the ECJ to remove trade barriers, do in fact cause greater trade (Pitarakis and Tridimas, 2003a). We address this and the more general issue of endogeneity in Section 3 and 4 below. The other reason to doubt the trade story is that the scope of EU law has expanded far beyond its initial focus on trade. This evolution is reflected in the sharp decline in the share of preliminary references pertaining to free movements of goods provisions (Figure 4).

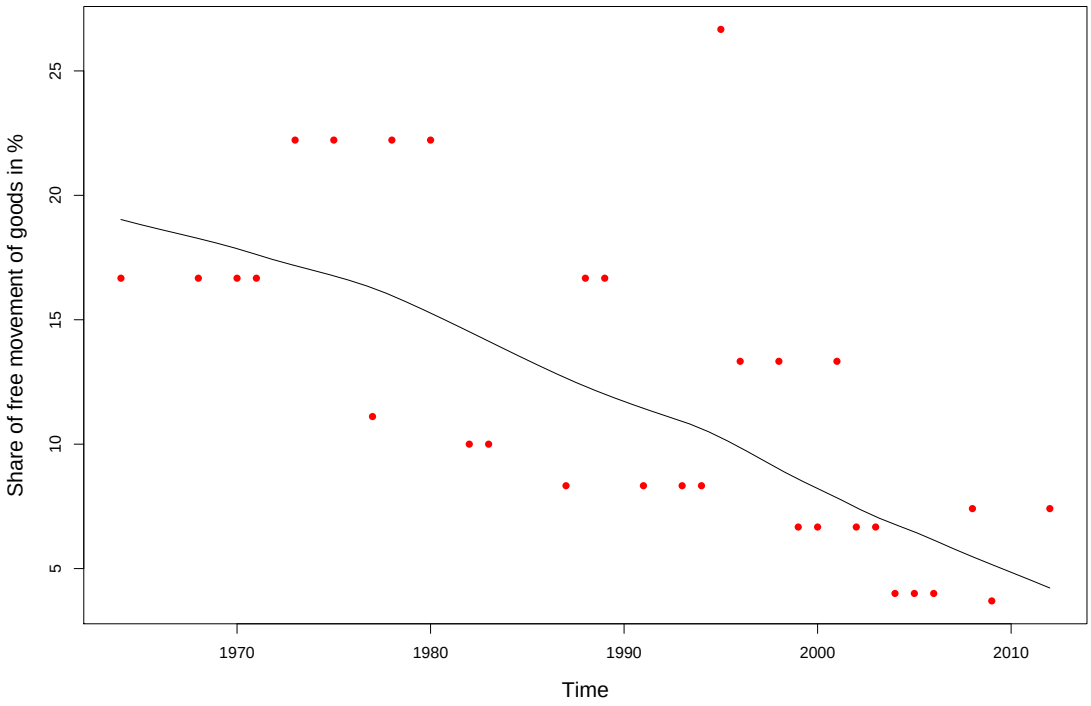


Figure 4: Share of preliminary references pertaining to free movements of goods provisions

It is clear that most legal disputes giving rise to a preliminary reference do not revolve around trade. We hypothesise, therefore, that trade dynamics will not influence the judges' referral behaviour:

> $Hypothesis_5$: Intra-EU trade will not influence use of the preliminary ruling mechanism.

While EU law questions are still more likely to surface in some policy areas than in others, references engage a diverse and wide-ranging set of legal domains (see Appendix). Again, due to the limitations of existing indicators, we cannot test how the substance of domestic patterns of litigation affects the likelihood that judges will use the preliminary ruling system. However, we expect that larger member states with a higher litigation volume will generate more references. This can be viewed as a mere application of the law of large numbers. As more disputes are litigated by a larger number of judicial officers, the set of cases where EU law issues are likely to be raised and where, for the judge, the benefits of obtaining a preliminary ruling exceed its costs is bound to grow larger as well:

> $Hypothesis_6$: Higher litigation volumes will result in higher number of references.

2.3 Macro-Level Constraints: Public Support and Political Fragmentation

Macro-level constraints reflect the fact that the judicial branch is itself embedded in a political system. A vast and well documented body of scholarship emphasise the importance of political fragmentation and public support in determining the judiciary's level of political discretion (Clark, 2009; Vanberg, 2005; Dyevre, 2010; Tsebelis, 2002a; Gibson et al., 1998). Some insights from this strand of judicial behaviour research do seem to have found an echo in the work of students of the preliminary ruling system. (Carrubba and Murrah, 2005a, 405) invoke research by Gregory Caldeira on public support to derive and test the proposition that greater Euroscepticism discourages Member States judges from submitting references. Also, Wind et al. (2009b) argue (albeit without reference to previous judicial politics research) that majoritarian democracy works to inhibit use of the preliminary reference mechanism. Yet we discern several flaws in the way these insights have been applied to the EU context. For a start, research on judicial legitimacy and public support draws a clear distinction between diffuse support–support for the institution–and specific support–support for particular decisions (Gibson et al., 1998). Strong diffuse support means that courts can afford to issue decisions that have low specific support, as long as they do not do so across all the cases they adjudicate. So, to undermine diffuse support, the relatively rare decision to send a reference to the Court of Justice would not only have to be regarded as controversial but the event itself would have to be sufficiently salient for large segments of the public to be aware of it (Vanberg, 2005). Moreover, even if the reference eventually results in a controversial policy outcome, the ensuing public outcry trigger is more likely to be directed at the Court of Justice than at the domestic court implementing its preliminary ruling. For these reasons, we do not expect public support for EU membership to have an influence on referral behaviour:

> $Hypothesis_7$: Public support for EU membership will not affect propensity to submit references.[8]

As for political fragmentation, its operationalization in Wind et al. (2009a) lacks clear theoretical grounding. Denmark, Sweden, Finland and the UK are classified as majoritarian democracies because their political system is viewed as committed to the principle of parliamentary sovereignty (Wind et al., 2009b, 72). But what the paper leaves out is the reason why other Member States, such as the Netherlands, whose Constitution appears committed to the same principle, do not qualify as majoritarian democracy. So, instead of drawing on Wind et al. (2009b)'s dubious distinction between majoritarian and constitutional democracy, we adopt a definition of political fragmentation more consonant with broader judicial behaviour research.

[8]Note that our expectation would differ if the explanatory variable was diffuse support for the judiciary rather than specific support for EU membership. Given that public opinion surveys of attitudes towards the judicial branch are few and far between, however, we cannot subject this hypothesis to systematic empirical testing.

We operationalize it as a function of the number and ideological distance among veto players in the political system. More numerous and more polarized veto players afford judges more leeway to pursue their policy agenda without fear of legislative override or other court-curbing measures. We expect that this will make domestic courts keener to use the preliminary ruling system:

> $Hypothesis_8$: More numerous veto players and and greater ideological distance among them will result in greater propensity to submit references.

3 Data

To test the hypotheses set forth in the previous section, we extend the dataset originally coded by Sweet and Brunell (1998a) to include all Article 267 references from 2006 to 2014.[9] Similar to Stone Sweet and Brunell, we collected information on year, case number, submitting court and Member State of origin, adding more than 2900 references to the original dataset.

We use population as proxy for litigation volume, on the assumption that more populous Member States have larger judiciaries processing disputes in larger numbers. We measure transnational economic activity as the sum of Intra-EU exports and imports, using European Commission data (2016). As mentioned above and discussed at length in Pitarakis and Tridimas (2003b), assessing the relationship between trade and preliminary references requires addressing reverse causality as well as the possibility of omitted variable bias. If legal integration is itself endogenous to transnational economic activity, ordinary least square is likely to produce biased estimates. Instrumental variable (IV) estimation is a common strategy to address this challenge and purge the predictor variable of endogeneity (Greene, 1997; Wooldridge, 2010). A reasonable instrument for the problem that concerns us here is Foreign Direct Investment (FDI). Empirical studies show FDI to be positively correlated with intra-EU trade (Pelkmans, 2006; Denisia, 2010; Ambroziak, 2010). A large macroeconomic literature has also shown that FDI stimulates employment, higher productivity, technological spillovers while fostering higher exports and access to international markets (Denisia, 2010). Specifically,Ambroziak (2010) highlights FDI's positive impact on intra-industry trade in eight new Member States over the period 1995-2007. At the same time, though, we do not expect FDI to have an effect on preliminary references. The only way FDI can be correlated with references is through its effect on cross-border trade. FDI thus satisfies the two conditions for instrument selection. We use FDI data compiled by Eurostat (2013) for all EU-28 for the period 1961-2014.

In keeping with previous authors, we rely on Eurobarometer data to measure citizens' perceptions of European integration and test $Hypothesis_7$.[10]. Following Carrubba and Murrah (2005a),

[9] (Sweet and Brunell, 1998a)'s original dataset extended to 2006 is available at.

[10] Eurobarometer data is freely available from GESIS Leibniz Institute for the Social Sciences, see for further details: http://www.gesis.org/eurobarometer-data-service/home/.

we define public support for EU integration as the net difference in the share of respondents saying that EU-membership is a “good thing" and that saying it is a “bad thing".

To test $Hypothesis_8$, we use the veto player index developed and compiled by Detlef (2010) for 23 OECD countries from 1944 to 2012 (Detlef et al., 2014). The veto player index is constructed from information on the number, coherence and ideological distance among veto players in the political system (Tsebelis, 2002b; Detlef et al., 2014). Because this data does not cover all EU-28 Member States, we also draw on the World Bank Database of Political Institutions (DPI), which features a range of indicators on checks and balances and political veto points (Keefer, 2012). Because operationlization It bears emphasis, though, we regard the latter measure as less accurate than the former.[11]

We employ data from the Comparative Constitution Project (CCP) (Elkins et al., 2014) to determine which Member States have set up a specialized constitutional court and to establish the existence and scope of judicial review for each country-year. To measure the extent to which a Member State's legal tradition is committed to monism, we use a dummy variable from CCP (Elkins et al., 2014) that takes the value one if the constitution grants international treaties priority over ordinary legislation and the value zero otherwise. While this measure, arguably, does not capture the whole gamut of doctrinal variations in the domestic reception of international law, it has for itself that is well-specified and fully replicable.[12]

Rather than relying on proxy measures such as the number of students attending the College of Bruges (Hornuf and Voigt, 2015), we simply measure exposure to EU law by the number of years since accession. The longer EU law has been around, the more likely it is that domestic judges will come in contact with EU law arguments either through colleagues or litigants. Finally, we control for enlargement effects, which so far have only been discussed by Wind et al. (2009a) by creating eight dummy variables for each enlargement wave: founding members (BE, DE, FR, IT, LU, NL), 1973 (DK, IR, UK), 1981 (EL), 1986 (ES, PT), 1995 (AT, FI, SE), 2004 (CY, CZ, EE, HU LV, LI, MT, PL, SK, SL), 2007 (BG, RO) and 2013 (HR).

[11] Correlation between the two measures is $r = 0.25$, suggesting large discrepancies in variable operationalization and/or coding accuracy.

[12] As alternative measure we considered two indicators evolved by Mila Versteeg and Pierre Verdier which capture whether, respectively, ratified treaties and customary international law require implementing legislation to become part of domestic law. But the temporal and geographical coverage of these indicators is restricted. Results based on these alternative measures are provided in the Appendix (see table 5).

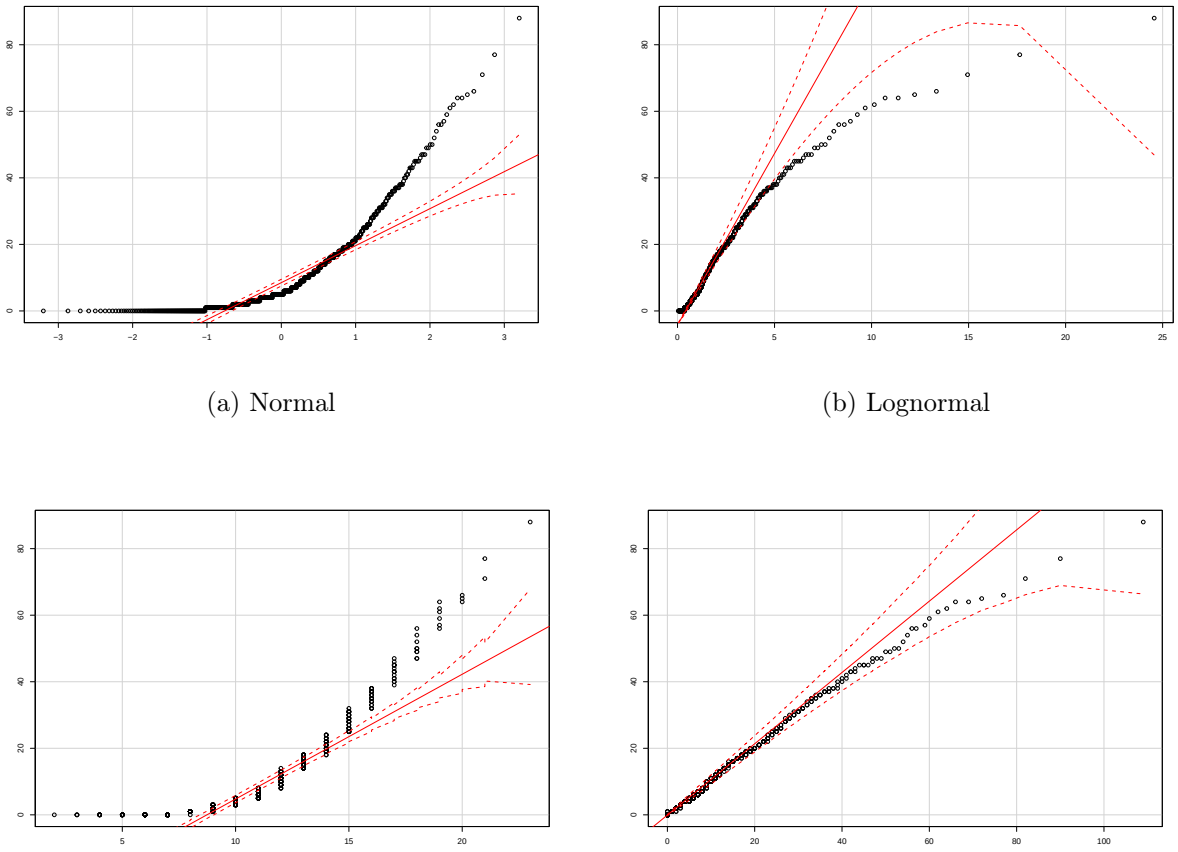
4 Estimation Strategy

In these section, we present our Bayesian framework along with our priors and treatment of endogeneity via IV estimation.

4.1 Dependent Variable: Distributional Form

Our dependent variable is defined as the absolute number of references submitted by a Member State's judiciary in a year. Thus defined the variable can only take on non-negative integer values. It has no upper bound and its distribution is positively skewed (Wooldridge, 2010). Some scholars have pointed out the problems raised by the count nature of this variable (Carrubba, 2005; Hornuf and Voigt, 2015; Kelemen and Pavone, 2016). But others have opted for functional forms that do not guarantee positivity for any parameter value (Sweet and Brunell, 1998b; Fligstein and Stone Sweet, 2002; Wind et al., 2009a; Gabel and Carubba, 2009). The reason this particular question warrants careful attention is that count data bounded at zero may invalidate linear model assumptions. An appropriate model for preliminary reference counts should not generate negative predicted values. Figure (5) compares four "plausible"[13] probability distributions by plotting empirical quantiles of the dependent variable[14] (horizontal axis) against theoretical quantiles (vertical axis). Clearly, the best fit is provided by the negative binomial probability distribution (panel 5d), as we see that all values fall within the confidence band (outer dashed line). Other distributional forms, by contrast, poorly fit the empirical distribution of reference counts.

[13] By "plausible" we mean either employed in the literature or plausible fit for count data.
[14] Furthermore, the frequency of the number of referrals is presented in the appendix (Figure 15f.)



(a) Normal

(b) Lognormal

(c) Poisson

(d) Negative binomial

Figure 5: Comparison of multiple probability distributions by plotting empirical quantiles of the dependent variable against theoretical quantiles

In our Bayesian estimation we use parameters from the fitted negative binomial distribution as our priors.

4.2 Bayesian Framework

The generic goal of our modelling approach is to explain cross-national referral dynamics over time conditional on specific factors. An appropriate model for such data must first reflect its longitudinal nature, i.e. the fact that we are dealing with repeated observations of the same Member States over time. Mixed models fit this requirement.[15] Flexibility and versatility make this class of models eminently well-suited for unbalanced and complex data structures (Galwey, 2006; **?**). Following the logic of mixed models, our model specifications include both a fixed (systematic) and random effects (unsystematic) part. While our substantive focus is on the

[15] Mixed models are also known in the literature as hierarchical or multi-level models.

fixed, systematic part of the cross-national variation, the random effect part serves to account for idiosyncratic variations among Member States and the unobserved factors that drive them. Technically, this is achieved by the inclusion of a random intercept for country-specific variations.

Assuming a non-negative, unbounded dependent variable Y_{ij} for the i^{th} observation in the j^{th} country and for n_j respondents, we can express the equation of interest as:

$$E(Y_{ij}|x_i, \nu_i, c_i) = exp(\beta x_{ij} + \delta \nu_{ij} + \alpha c_{ij} + \epsilon_{ij}) \tag{1}$$

where x_{ij} are the fixed-effects covariates; β the fixed effects coefficient; c_{ij} random-effects covariates; α the random effect coefficient; ν_{ij} is an observable explanatory variable (δ the estimation coefficient) and ϵ_{ij} is the noise term. Let us suppose that ν_{ij} is correlated with ϵ_{ij}. Reverse causality and the possibility that omitted variables affect both transnational economic activity and use of the preliminary ruling system entail that ν_{ij} is potentially endogenous. To purge the endogenous regressor ν_{ij} from endogeneity, we apply a two stage least squares (2SLS) estimation procedure in combination with a control function approach (Greene, 1997; Wooldridge, 2010). This allows us to estimate our endogeneous variable as a linear function of control variables and instrument (i.e. FDI). The equation of the second step (reduced form equation) can be written as:

$$\nu_{ij} = \gamma x_{ij} + \rho z_{ij} + \upsilon ij \tag{2}$$

where x_{ij} are the control variables and γ the estimated coefficient: z_{ij} is the instrumental variable and ρ the estimated coefficient and υ_{ij} is the reduced form error. Wooldridge (2010) proposes to estimate both equations under more restrictive assumptions (control function approach) when the structural model of interest (equation 1) is of exponential form. To that end, we posit that the dependent variable y given the instrument z and the endogenous variable ν given ϵ follow a negative binomial distribution with exponential mean.

We estimate our main equation of interest (equation 1) through Bayesian inference in a hierarchical set-up. Owing to its flexibility and versatility, Bayesian hierarchical modelling performs well with complex structures (multiple levels) as well as with sparse and unbalanced longitudinal data (Gill, 2002; **?**; **?**). "Bayesian learning", obviously, is one of the characteristics that makes this approach attractive for the analysis of longitudinal or time series data. Our model treats previous posteriors as new priors and updates them in light of each additional observation.

We set up the hierarchical model to express within-variation at the country level and between-variation at time level. Adapting the notation and re-writing 1 in Bayesian terms, let Y_{ij} be the number of preliminary references observed in period i in Member State j with $i = 1, \ldots, n$ and $j = 1, \ldots, J$. Let $x_{ij} = [1, x_{ij}^1, \ldots, x_{ij}^P]^K$ be the covariate vector (fixed effects)[16], and let

[16]In hierarchical Bayesian inference, the notions of fixed and random effects are formally undistinguishable as

$\beta_j = [\beta_j^0, \ldots, \beta_j^P]^K$ be the unknown estimation coefficient and ψ_j be the dispersion parameter of the corresponding negative binomial (NB) distribution.

Given that $(x_{ij}, \beta_j, \psi_j)$, Y_{ij} is expected to follow the NB distribution[17] which can be expressed as:

$$
\begin{aligned}
Y_{it}|p_{ij} &\sim \mathcal{NB}(p_{ij,\psi_j}) \\
pij &= exp(\underset{(1\times k)}{X_{ij}} \quad \underset{(k\times 1)}{\beta_j}) \\
\beta_{jk} &= \underset{1\times l}{G_{kl}} \; \underset{l\times 1}{\eta_k} + \alpha_{jk} \\
\alpha_{jk} &= t(\nu, 0, s) \\
\eta &= \mathcal{N}(m_n \Sigma_n)
\end{aligned}
$$

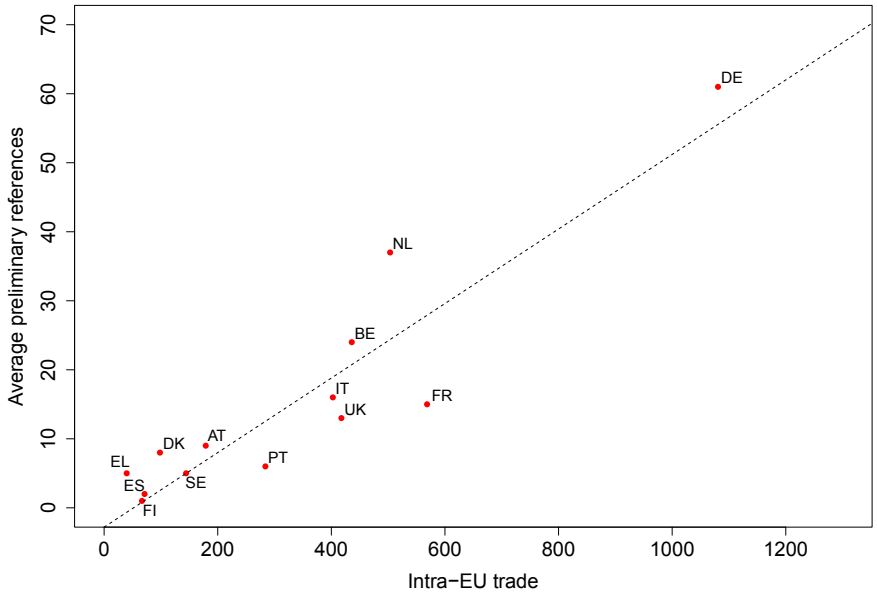
where β_j is the k^{th} country coefficient vector and each of the k country coefficients b_{jk} is explained by a higher-level parametrization (intercept and slope vary across countries) in which G_{kl} is a vector of second level explanatory variables (variation across the periods) and η_k is the underlying second-level coefficient. The residual error term α_{jk} follows a t-student prior with $\nu = 3$ and scale $s = 10$ and we set a flat prior on η_k. To demonstrate the robustness of our findings, we run our model with both proper and improper priors. The results are reported in the next section.

5 Empirical results

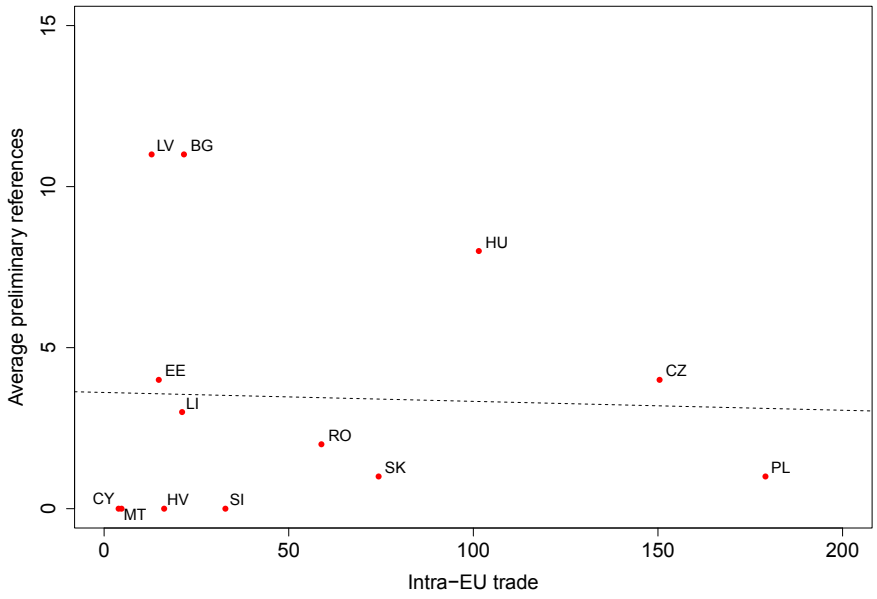
Because trade has featured so prominently in the scholarly discussion, we first briefly consider characteristics of the data that initially led researchers to believe transnational economic activity and referral behaviour are causally linked. Plotted in panel (a) in Figure 6 are average annual free movements of goods reference counts against intra-EU trade to GDP ratio for the EU-15 Member States for the period 2006-2014.

all parameters are considered random.

[17] The probability density function of the negative binomial distribution can be written as: $f(x_{ij}, \beta_j, \psi_j) = \frac{\Gamma(y_{ij}+\psi_j)}{y_{ij}!\Gamma(\psi_j)} \left[\frac{p_{ij}}{p_{ij}+\psi_j}\right]^{y_{ij}} \left[\frac{psi_j}{p_{ij}+\psi_j}\right]^{\psi_j}$.



(a) Old member states



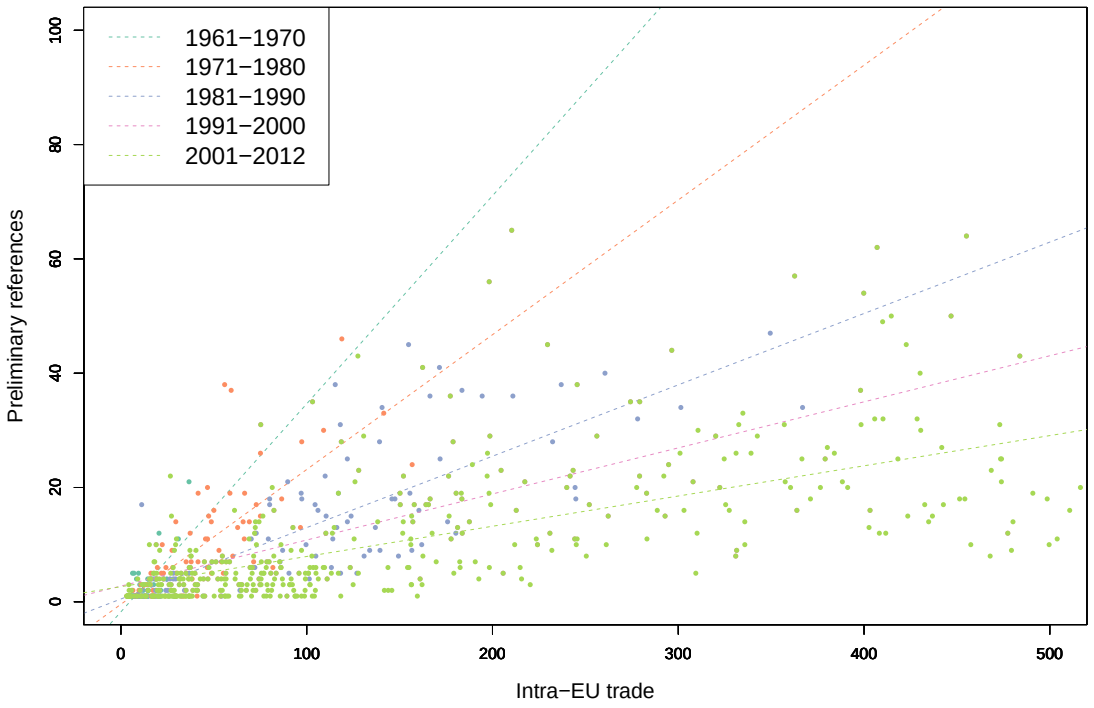
(b) New member states

Figure 6: Intra-EU Trade and Average Annual Reference Rates, Free Movement of Goods Cases, 2006-2014

The strong, positive relationship between trade and free movement of goods references for this group of Member States is virtually identical to the results plotted by Sweet and Brunell

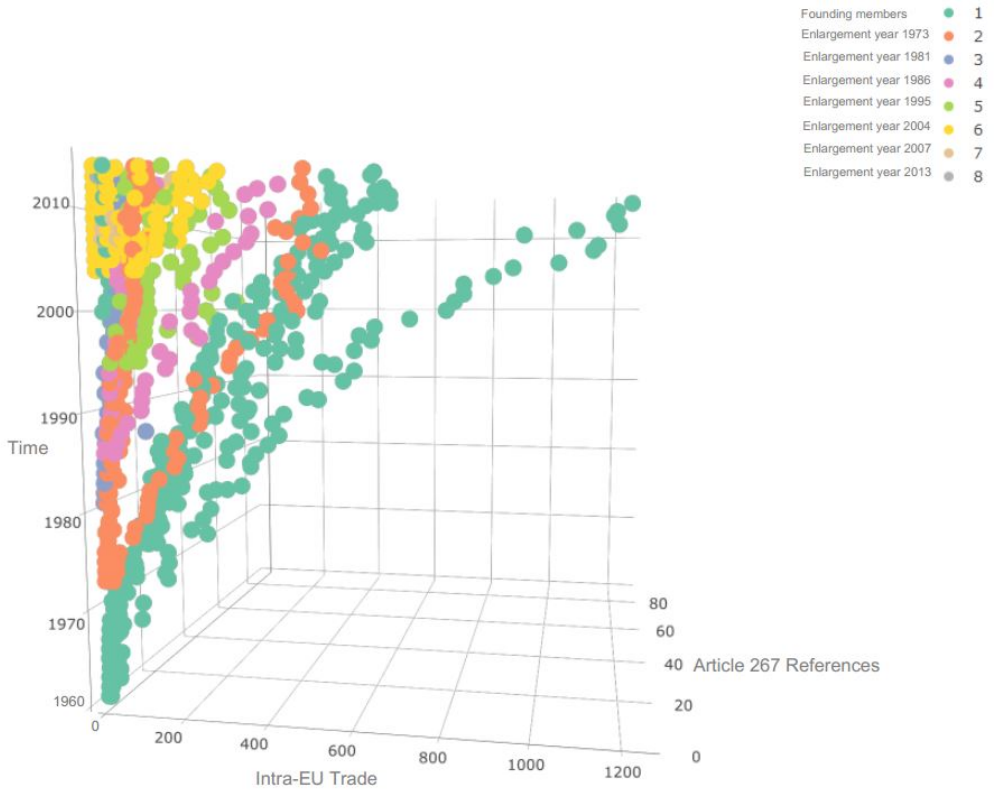
(1998b). Looking at new Member States, however, the picture turns out to be very different. For this set of Member States the relationship between trade and references is negative. This result, though, is not necessarily inconsistent with the theory advanced by (Sweet and Brunell, 1998b). Indeed, Stone Sweet and Brunell did theorise that the relationship between trade and preliminary references would weaken over time, as the scope of EU law expanded and litigation spilled over to other areas (Sweet and Brunell, 1998b). At first glance, this explanation seems to make good sense. If we run separate regressions for each decade since 1961, we see that the slope of the regression declines steadily (Figure), in pace with the shrinking share of free movement of goods references.

Figure 7: Regression Coefficient, Average Annual Referral Rate and Intra-EU Trade



Yet the 3D plot in Figure 8 suggests a different story. What Figure 8 shows is that the correlation between trade and references was and remains strong for some founding Member States–Germany essentially–but that there is no clear pattern of relationship for most Member States.

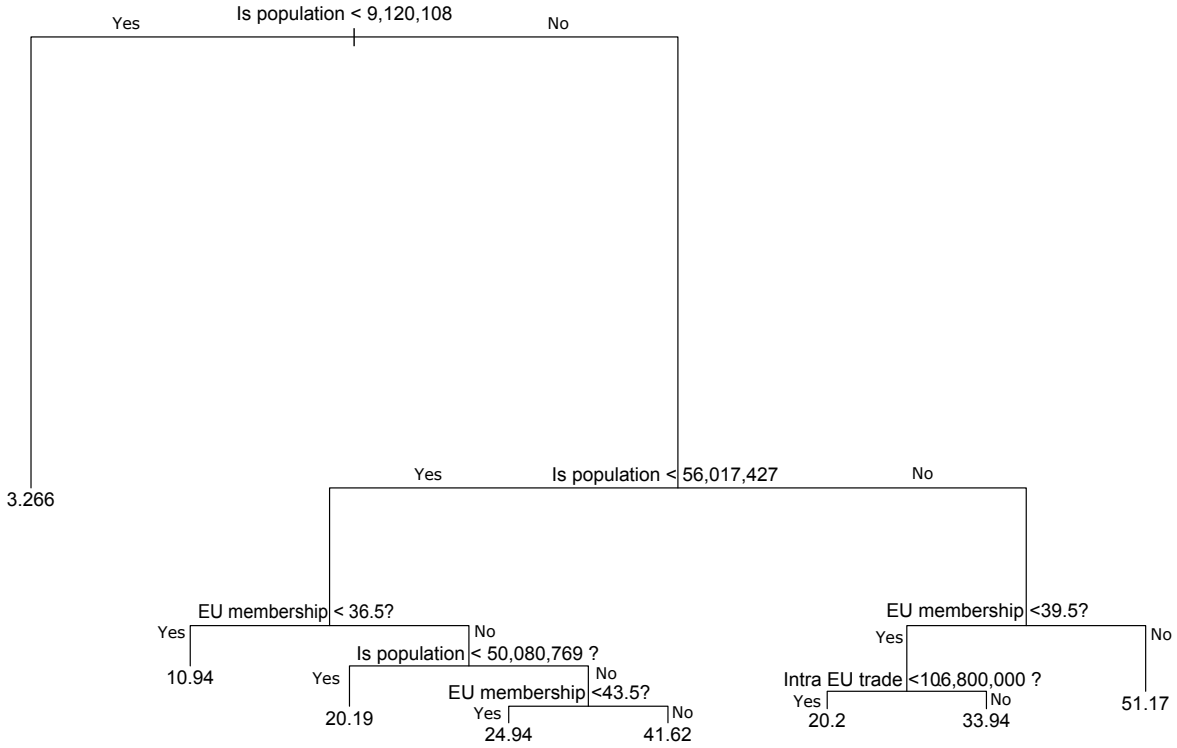
Figure 8: Relationship between Intra-EU trade and preliminary references over years



To explore non-linear relations between our dependent variable and covariates, we construct a regression tree.[18] Depicted in Figure is the resulting classification tree. The tree clearly suggests that population and years of membership are the strongest predictors of referral rates. Intra-EU trade does help classify Member States but only for the most populous Member States before they reach 39.5 years of membership. In fact, the criterion identified by the procedure (intra-EU trade) mostly serves to distinguish Germany from France and the UK prior to the year 1999 (beyond that mark intra-EU trade becomes irrelevant to the classification). This reinforces, again, our contention that the relationship between intra-EU trade and references is driven by Germany.

[18] Classification and Regression Tree (CART) models are usually deployed in the context of forecasting, but the point we make here is principally exploratory (Varian, 2014).

Figure 9: Regression Tree with Random Effects



Unlike our hierarchical Bayesian model, however, CART models of this sort do not account for endogeneity in the data. So we now present the results of our Bayesian estimation. To obtain our estimates we run 24000 stationary iterations, burning the first 8000 as warm up iterates. We report results both with and without controlling for endogeneity. Due to gaps in some of the datasets, we report the results of different model specifications that reflect varying data availability. Presented in column 2-6 in table 2 are our estimation results for the EU-15 Member States (*Ia* and *Ib*). Column 7-13 report results and model specification for all EU-28 countries (*IIa* and *IIb*). The results presented here are based on our proper, empirical priors. As a robustness check, we also run the model with the same specifications using non-informative priors. Substantively, they differ little for the EU-15 Member States group from those presented here. However, the chains do not converge properly for the EU-28 Member States which indicates that there is a substantial part of noise in the data and thus stronger priors need to be specified for this group.

Density, trace and autocorrelation plots used to evaluate validity of the Bayesian Monte Carlo Markov Chain estimation process are also reported in the Appendix D. Reported here are the predictive accuracy of model fit and the Heidelberg-Welch autocorrelation tests.[19]

Mean posterior estimates are significant and positive for the covariates *Population* and *Years of membership* across all model specifications. Consistent with the results from the regression tree above, these results lend support to $Hypothesis_1$ and $Hypothesis_6$.

Result 1: *Higher exposure to EU law and higher litigation volumes are affecting positively and significantly the use of preliminary mechanism*

Estimates for intra-EU trade are negative and significant when we do not control for endogeneity (*Ia* and *IIa*). Computing the same coefficients via IV estimation, though, yields posterior means that are no longer significant (*Ib*) and (*IIb*).[20] While consistent with Wind et al. (2009a) and Gabel and Carubba (2009), these results diverge from Fligstein and Stone Sweet (2002), Carrubba and Murrah (2005b) and Gabel et al. (2012). Yet none of these studies addressed the issue of endogeneity in the relationship between trade and referral rates.

Result 2: *Intra-EU trade has no influence on use of the preliminary ruling mechanism*

Posterior mean estimates for political fragmentation are positive and significant for all model specifications for EU-15 Member States. On the other hand, when we use DPI instead of veto player index data (the latter is only available for EU-15) and run the model for all 28 Member States, posterior mean estimates come out as non-significant, albeit still positive(model (*IIa*) and (*IIb*). Because we believe the veto player index provides a more reliable measure of political fragmentation, we are still confident enough to conclude that our results support $Hypothesis_8$.

Result 3: *Greater political fragmentation makes judges keener to submit references*

As we hypothesised, support for EU membership is not significant, whether we only consider the EU-15 and EU-28.

Result 4: *Public attitudes towards European integration has no systematic effect on referral rates*

[19] A critical part of the Bayesian estimation process, using Markov Chain Monte Carlo (MCMC) method, is to assess the model fit to the data and report the sensitivity of the posterior distribution to the assumptions (Cowles and Carlin, 1996; Gill, 2002). There are various ways to examine model integrity by numerical or visual diagnostics. Gelman et al. (2014) provides helpful guidelines along with examples on how to investigate an accurate assessment of Bayesian models. See also Cowles and Carlin (1996) for a comparative review about the application of convergence diagnostic tools.

[20] We run the weak instrument and Durbin-Wu-Hausman tests to reject the null hypothesis at 1% significant level that transnational economic activity is exogenous.

That referral propensity is not influenced by specific support of EU membership is consistent with mainstream judicial behaviour research, which emphasises the importance of diffuse, rather than specific, support. Still, while consistent with theory, this finding appears remarkable when considered against the dramatic decline in public support for EU membership over the past two decades (10). Neither the financial crisis, nor the Euro crisis have had a systematic impact on the rate at which domestic courts refer questions to the Court of Justice.

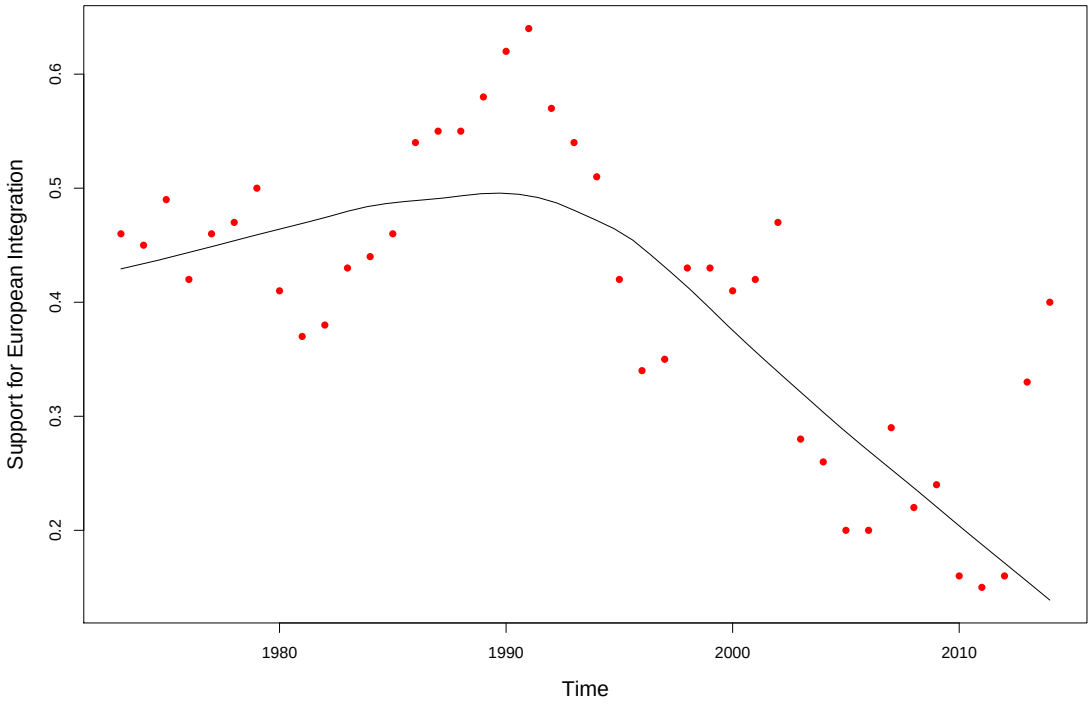


Figure 10: Public Support for EU Membership

Legal covariates–monism, judicial review, constitutional court–come out as not significant in most model specifications. The mean posterior estimate for *Judicial review* is negative and significant the EU-15 (*Ia* and *Ib*) but the 95% confidence interval for the EU-28 does include zero (albeit not by much). In sum, we find weak evidence for the empowerment hypothesis ($Hypothesis_3$) but no evidence for the court competition hypothesis ($Hypothesis_4$), while results for the variable *Monism* confirm $Hypothesis_2$.

> **Result 4:** *Greater powers to exert judicial review under domestic law result in lower propensity to use the preliminary ruling mechanism. However, monism and the existence of a constitutional have no significant effect on referral behaviour*

Turning now to the impact of enlargement, our results suggest that none of the enlargement

waves has had a consistent, cross-national effect on the preliminary ruling mechanism except for the fourth wave (1995), when Austria, Finland and Sweden joined the EU (*Ia*, *IIa* and *IIb*).[21] The impact of the fourth wave can largely be attributed to the enthusiasm of Austrian courts for the preliminary ruling system.

> **Result 5:** *Save for the fourth enlargement wave, successive enlargements have had no impact on referral rates*

These findings run counter to Wind et al. (2009a) who observe differing patterns of referral behaviour associated with distinct enlargement waves, including 1973 and 1981.

[21] The reference category for our categorical variables is the group of founding Member States (Belgium, France Germany, Italy, Luxembourg and Netherlands). It is captured by the posterior mean of the constant.

Table 2: Hierarchical Bayesian model estimation

	EU15 Member States (*I*)						EU28 Member States (*II*)					
	Without Endogeneity (*Ia*)			With Endogeneity (*Ib*)			Without Endogeneity (*IIa*)			With Endogeneity (*IIb*)		
Var	Post.mean	l-95% CI	u-95% CI	Post.mean	l-95% CI	u-95% CI	Post.mean	l-95% CI	u-95% CI	Post.mean	l-95% CI	u-95% CI
Constant	-4.83	-6.647	-2.798	-3.894	-5.561	-2.060	-3.531	-5.461	-1.546	-2.757	-4.592	-0.828
Trade.gdp	-0.019	-0.036	-0.002	-0.005	-0.029	0.019	-0.019	-0.038	-0.001	-0.018	-0.056	0.020
Population (log10)	1.473	1.038	1.837	1.304	0.899	1.657	1.120	0.685	1.532	1.003	0.578	1.401
Years of membership	0.04	0.027	0.053	0.03	0.021	0.038	0.038	0.025	0.051	0.029	0.020	0.037
Political fragmentation (Veto)	0.013	0.002	0.024	0.012	0.001	0.023						
Political fragmentation (Checks)							0.037	-0.051	0.124	0.036	-0.055	0.126
EU support	0.132	-0.372	0.627	0.313	-0.154	0.787	-0.061	-0.542	0.419	0.112	-0.354	0.578
Judicial Review	-0.149	-0.275	-0.03	-0.147	-0.277	-0.023	-0.080	-0.221	0.057	-0.084	-0.227	0.053
Constitution	-0.165	-0.642	0.259	-0.172	-0.656	0.28	0.006	-0.598	0.584	-0.004	-0.599	0.568
Monism	-0.604	-1.114	0.069	-0.515	-1.075	0.162	-0.070	-0.663	0.546	-0.061	-0.647	0.570
Enlargement year 1973	-0.378	-0.929	0.166	-0.474	-1.052	0.134	-0.461	-1.290	0.372	-0.526	-1.321	0.300
Enlargement year 1981	0.855	-0.033	1.58	0.704	-0.184	1.487	0.155	-0.879	1.202	0.122	-0.908	1.144
Enlargement year 1986	0.267	-0.312	0.887	0.066	-0.514	0.698	0.252	-0.593	1.126	0.067	-0.763	0.925
Enlargement year 1995	0.808	0.074	1.543	0.593	-0.124	1.307	0.938	0.086	1.794	0.766	-0.083	1.606
Enlargement year 2004							-0.179	-1.408	1.058	-0.288	-1.511	0.936
Enlargement year 2007							0.596	-0.655	1.874	0.528	-0.739	1.794
NMS							0.394	-0.869	1.641	0.236	-0.989	1.493
σ_i		0.23			0.25			0.48			0.47	
WAIC		1271.00			1276.51			1576.83			1581.28	
$R^2_{glmm(m)}$		0.595			0.451			0.421			0.316	
$R^2_{glmm(c)}$		0.791			0.717			0.708			0.654	
Heidelberger and Welch Stationary test (σ_i)		$p_{\text{value}} = 0.114$			$p_{\text{value}} = 0.110$			$p_{\text{value}} = 0.474$			$p_{\text{value}} = 0.340$	
Heidelberger and Welch Halfwidth interval test (σ_i)		Mean= 0.221 Halfwidth= 0.005			Mean= 0.234 Halfwidth= 0.005			Mean= −0.079 Halfwidth= 0.001			Mean= −0.084 Halfwidth= 0.001	
Number of iterations	16000						16000					
Number of country levels	13						24					
Original sample size	222						298					
Weak Instrument test	F-statistic = 18.179***						F-statistic = 61.914***					
Hausman-Wu test	F-statistic = 9.204***						F-statistic = 19.012***					

Note: Informative priors are specified on the basis of available information over parameter values and correspond to a negative-binomial distribution for response variable;

Constant includes the reference category *EU Founding Member States* (Belgium, Germany, France, Luxembourg, Italy and Netherlands).

As for the overall variance explained by our model, computing the equivalent of a R^2 is not as straightforward as for linear models due to the presence of error terms at several levels. Following Nakagawa and Schielzeth (2013) we report two measures of explained variance for mixed effects model–marginal ($R^2_{glmm(c)}$) and conditional R^2. The former captures the variance explained by fixed effects, while the latter captures the variance explained by both fixed and random effects. Based on conditional $R^2_{glmm(c)}$ and marginal $R^2_{glmm(m)}$ and Bayesian information criterio (WAIC), model specifications (Ia) and (IIa) exhibit the best goodness of fit and highest explained variance.

6 Conclusion

In this paper we tried to place the empirical study of the preliminary ruling system on more rigorous theoretical and methodological foundations. We presented an account of judicial motivation in the context of the preliminary ruling mechanism which, compared to accounts grounded in integration theory, has the advantage of greater consilience. Our empirical findings are broadly consistent with the picture of judicial decision making that has emerged from research in other settings. We also believe that they are more robust than those yielded by previous studies.

This said, however, we think that there is room for more research in this area. Better and more sophisticated measures for some of the explanatory factors–litigation, legal doctrines and political fragmentation in the new Member States–would help refine our findings. Yet there are inherent limits as to what a model will be able to explain as long as we aggregate references at country-level. With the notable exception of Kelemen and Pavone (2016), all studies, including the present one, have treated national judiciaries as unitary agents. There are, however, great disparities not only across courts within the same Member State but also possibly across judicial panels within the same court that may explain a great deal of the referral dynamics we observe. Whether it is by re-defining the dependent at the regional level, as (Kelemen and Pavone, 2016) for references submitted by Italian courts, or at the court-level, we believe that future research should move to the infra-state level.

References

Alter, A. L. and D. M. Oppenheimer (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and social psychology review*.

Alter, K. J. (1998). Explaining National Courts Acceptance of European Court Jurisprudence: A Critical Evaluation of Theories of Legal Integration. In A.-M. Slaughter, A. Stone Sweet, and J. Weiler (Eds.), *The European Court and National Courts - Doctrine and Jurisprudence*, pp. 232.

Alter, K. J. (2001). *Establishing the supremacy of European law: The making of an international rule of law in Europe*. Oxford University Press Oxford.

Ambroziak, u. (2010). The foreign direct investments (fdi) as a factor of intra-industry trade development in the eu new member states. European Trade Study Group (ETSG).

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.

Burley, A.-M. and W. Mattli (1993). Europe Before the Court: A Political Theory of Legal Integration. *International Organization 47*(01), 41–76.

Carrubba, C. and L. Murrah (2005a). Legal integration and use of the preliminary ruling process in the european union. *International Organization 59*, 399–418.

Carrubba, C. J. (2005). Courts and Compliance in International Regulatory Regimes. *J. of Pol. 67*(3), 669–689.

Carrubba, C. J. and L. Murrah (2005b). Legal Integration and Use of the Preliminary Ruling Process in the European Union. *International Organization 59*(02), 399–418.

Chen, D., T. J. Moskowitz, and K. Shue (2016). Decision-Making Under the Gambler's Fallacy: Evidence from Asylum Judges, Loan Officers, and Baseball Umpires. *The Quarterly Journal of Economics*, qjw017.

Clark, T. S. (2009). The separation of powers, court curbing, and judicial legitimacy. *American Journal of Political Science 53*(4), 971–989.

Cowles, M. K. and B. P. Carlin (1996). Markov chain monte carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association 91*(434), 883–904.

Danziger, S., J. Levav, and L. Avnaim-Pesso (2011, April). Extraneous factors in judicial decisions. *PNAS 108*(17), 6889–6892.

Denisia, V. (2010). Foreign direct investment theories: An overview of the main fdi theories. *European Journal of Interdisciplinary Studies 2*(2), 104–110.

Detlef, J. (2010). *The Veto Player Approach in Macro-Comparative Politics: Concepts and Measurement. In Reform Processes and Policy Change. Veto Players and Decision-Making in Modern Democracies*, Chapter 3, pp. 43–68. New York, NY: Springer.

Detlef, J., T. Behm, N. Düpont, and C. Oberst (2014). Pip – parties, institutions & preferences: Veto player (annual) [version 2014-09]. Chair of Comparative Politics, University of Greifswald.

Downs, G. W., D. M. Rocke, and P. N. Barsoom (1996, June). Is the good news about compliance good news about cooperation? *International Organization 50*(03), 379–406.

Dyevre, A. (2010). Unifying the Field of Comparative Judicial Politics: Towards a General Theory of Judicial Behaviour. *European Political Science Review 2*(02), 297–327.

Elkins, Z., T. Ginsburg, and J. Melton (2014). Comparative constitutions project. characteristics of national comparative. version 2.

Epstein, L. and J. Knight (2013, May). Reconsidering Judicial Preferences. *Annual Review of Political Science 16*(1), 11–31.

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.

European Commission (2016). Annual macro-economic database (AMECO), asessed on 17 october 2016. irectorate General for Economic and Financial Affairs (DG ECFIN).

Eurostat (2013). Structural buisness statistics, assessed on 12 november 2016. http://ec.europa.eu/eurostat/web/structural-business-statistics/global-value-chains/fdi.

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.

Gabel, M., C. Carrubba, C. Ainsley, and D. Beaudette (2012). Of courts and commerce. *The Journal of Politics 74*(4), 1125–1137.

Gabel, M. and C. Carubba (2009). The european court of justice as an engine of economic integration: Reconsidering evidence that the ecj has expanded economic exchange in europe.

Galwey, N. (2006). *Introduction to Mixed Modelling. Beyond regression and analysis of variance.* Wiley.

Garrett, G. (1995). The politics of legal integration in the European Union. *International Organization 49*, 171–171.

Gelman, A., J. Hwang, and A. Vehtari (2014). Understanding predictive information criteria for bayesian models. *Statistics and Computing 24*, 997–1016.

Gibson, J. L., G. A. Caldeira, and V. A. Baird (1998, June). On the Legitimacy of National High Courts. *Am. Pol. Sci. Rev. 92*(2), 343.

Gill, J. (2002). *Bayesian methods. A Social and behavioral sciences approach.*

Greene, W. H. (1997). *Econometric Analysis* (3 ed.).

Hammond, T. H., C. W. Bonneau, and R. S. Sheehan (2005). *Strategic Behavior And Policy Choice On The U.S. Supreme Court.* Stanford University Press.

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.

Huneeus, A. (2013). Compliance with International Court Judgments Decisions. In C. Romano, K. J. Alter, and Y. Shany (Eds.), *The Oxford Handbook of International Adjudication*, pp. 438–463. Oxford University Press.

Jaremba, U. (2013). At the crossroads of national and european union law-experiences of national judges in a multi-level legal order. *Erasmus L. Rev. 6*, 191.

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.

Keohane, R. O., A. Moravcsik, and A.-M. Slaughter (2000). Legalized Dispute Resolution: Interstate and Transnational. *International Organization 54*(03), 457–488.

Mayoral, J. A., U. Jaremba, and T. Nowak (2014, June). Creating EU law judges: the role of generational differences, legal education and judicial career paths in national judges' assessment regarding EU law knowledge. *Journal of European Public Policy*.

Nakagawa, S. and H. Schielzeth (2013). A general and simple method for obtaining r2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution 4*, 133–142.

Pelkmans, J. (2006). *European integration: methods and economic analysis* (2 ed.). Harlow : Pearson.

Pitarakis, J. and G. Tridimas (2003a). Joint dynamics of legal and economic integration in the european union. *European Journal of Law and Economics 16*, 357–368.

Pitarakis, J.-Y. and G. Tridimas (2003b). Joint dynamics of legal and economic integration in the european union. *European Journal of Law and Economics 16*(3), 357–368.

Rogers, J. R., R. B. Flemming, and J. R. Bond (2006). *Institutional Games and the US Supreme Court*. University of Virginia Press.

Stephenson, M. C. (2009). Legal Realism for Economists. *The Journal of Economic Perspectives 23*(2), 191–211.

Stone Sweet, A. (2004). *The Judicial Construction of Europe*. Oxford [u.a.]: Oxford University Press.

Stone Sweet, A. and T. Brunell (2012). The European Court of Justice, State Noncompliance, and the Politics of Override. *American Political Science Review 106*(01), 204–213.

Sweet, A. and T. Brunell (1998a). The European Court and the national courts: a statistical analysis of preliminary references, 1961–95. *Journal of European Public Policy 5*(1), 66–97.

Sweet, A. S. and T. L. Brunell (1998b). 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.

Thagard, P. R. (1978). The best explanation: Criteria for theory choice. *The journal of philosophy 75*(2), 76–92.

Tsebelis, G. (2002a). *Veto Players: How Political Institutions Work*. Princeton University Press.

Tsebelis, G. (2002b). *Veto Players: How Political Institutions Work*.

Vanberg, G. (2005). *The Politics of Constitutional Review in Germany*. Cambridge [u.a.]: Cambridge Univ. Press.

Varian, H. R. (2014). Big data: New tricks for econometrics. *The Journal of Economic Perspectives 28*(2), 3–27.

Vink, M. and M. Claes (2008). Beyond pathology: Explaining the use of preliminary references by domestic courts in eu member sates. Paper for workshop 7 The Long aArm of EU Law: Making States Comply with EU Policies inside and outside the European Union.

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.

Weiler, J. H. H. (1994). A Quiet Revolution. *Comp. Pol. Studies 26*(4), 510–534.

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

Wind, M., D. S. Martinsen, and G. P. Rotger (2009b, March). The Uneven Legal Push for Europe Questioning Variation when National Courts go to Europe. *European Union Politics 10*(1), 63–88.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data* (Second ed.). The MIT Press.

A Explanatory tables

Table 3: Summary statistics of the meta-analysis

Author	Sample size	Time period	Intra-EU trade coefficient
Sweet and Brunell (1998a)	246	34	0.0001
Fligstein and Stone Sweet (2002)	576	36	0.0030
Pitarakis and Tridimas (2003b)	246	34	2.3630
Carrubba (2005)	270	28	0.3300
Wind et al. (2009a)	234	43	-0.2820
Vink et al. (2009)	176	11	-0,0980
Vink and Claes (2008)	176	11	7,7570
Gabel and Carubba (2009)	289	35	0.9500
Gabel et al. (2012)	120	23	0.8080
Hornuf and Voigt (2015)	331	38	1.0030
Kelemen and Pavone (2016)	272	49	-0.0350

Table 4: List of variables

Variable name	Explanation	Type of data	Source
$\text{Case}_{i,t}$ (Dependent variable)	Number of referrals per country i at time t	discrete	Sweet and Brunell (1999)
$\text{Trade}_{i,t}$	Intra EU-trade (export plus import) per GDP per capita per country i at time t	continuous	Ameco macro-economic database (European Commission, 2016)
$\text{Pop}_{i,t}$	Number of inhabitants per country i at time t	continuous	Ameco macro-economic database (European Commission, 2016)
Years of $\text{membership}_{i,t}$	Difference between the request date of preliminary reference at time t and the year the country i entered the European Union	continuous	NA
$\text{FDI}_{i,t}$	Foreign direct investment made by country i at time t.	continuous	Eurostat (2013)
EU $\text{support}_{i,t}$	Net political support per country i at time t; Percentage difference between respondents believing that EU membership is a good thing and those considering that EU membership is a bad thing.	continuous	Eurobarometer (2016)
Checks_i	is incremented by one if there is a chief executive, or if the chief executive is competitively elected or if the opposition controls the legislature	categorical	Keefer (2012)
$\text{Veto}_{i,t}$	Number of veto players per country i at time t	continuous	Detlef et al. (2014)
Judicial review_i	The review of constitutionality per country i is measured through an interaction term; The term is composed by the the stage of the legislative process at which a bill can be reviewed for constitutionality (chalstag) and the existence of judicial review	categorical	Comparative Constitution Project Elkins et al. (2014)
Monism_i	Dummy variable which takes the value one if the status of treaties is superior to ordinary legislation in the country i and zero otherwise	categorical	NA
Constitution_i	Dummy variable which takes the value one if country i has adopted Austrian model and zero otherwise	categorical	NA
Treaty Status_i	Dummy variable which takes the value one if ratified treaties automatically become part of domestic law without implementing legislation per country i and zero otherwise	categorical	NA
CIL_i	Dummy variable which takes the value one if Customary international law automatically becomes part of domestic law without implementing legislation per country i and zero otherwise	categorical	NA
Founding members_i	Dummy variable which takes the value one if the Member State i is in the founding group (BE, NL, LU, FR, IT, GE) and zero otherwise	categorical	NA
Enlargement year 1973_i	Dummy variable which takes the value one if the Member State i is in the enlargement group (UK, IR, DK) and zero otherwise	categorical	NA
Enlargement year 1981_i	Dummy variable which takes the value one if the Member State i is in the enlargement group (EL) and zero otherwise	categorical	NA
Enlargement year 1986_i	Dummy variable which takes the value one if the Member State i is in the enlargement group (ES, PT) and zero otherwise	categorical	NA
Enlargement year 1995_i	Dummy variable which takes the value one if the Member State i is in the enlargement group (AT, FI, SE) and zero otherwise	categorical	NA
Enlargement year 2004_i	Dummy variable which takes the value one if the Member State i is in the enlargement group (CY, EE, HU, LT, LI, MT, PL, CZ, SK, SL, BG, RO) and zero otherwise	categorical	NA
NMS_i	Dummy variable which takes the value one if the country i refers to new member state and zero otherwise	categorical	NA

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

B Alternative model specification

Table 5: Hierarchical Bayesian estimation using alternative measures

	EU15 Member States						EU28 Member States					
	Without Endogeneity (Ic)			With Endogeneity (Id)			Without Endogeneity (IIc)			With Endogeneity (IId)		
Var	Post.mean	l-95% CI	u-95% CI	Post.mean	l-95% CI	u-95% CI	Post.mean	l-95% CI	u-95% CI	Post.mean	l-95% CI	u-95% CI
Constant	-3.939	-7.624	0.203	-3.772	-7.426	0.141	-4.913	-8.296	-1.324	-4.653	-7.919	-1.264
Trade.gdp	-0.009	-0.028	0.009	0.007	-0.011	0.024	-0.009	-0.027	0.008	0.002	-0.021	0.024
Population (log10)	1.157	0.465	1.671	1.124	0.509	1.592	1.077	0.534	1.563	1.048	0.540	1.508
Years of membership	0.035	0.021	0.049	0.027	0.017	0.036	0.033	0.020	0.047	0.027	0.017	0.037
Political fragmentation (Veto)	0.012	0.001	0.022	0.012	0.002	0.023						
Political fragmentation (Checks)						0.095	0.006	0.186	0.087	-0.003	0.177	
Political support	0.539	0.01	1.068	0.614	0.13	1.103	0.429	-0.067	0.915	0.516	0.063	0.977
Judicial review	-0.162	-0.318	0.003	-0.141	-0.307	0.037	-0.070	-0.204	0.064	-0.063	-0.201	0.074
Ratified Treaties	-0.213	-1.061	0.543	-0.148	-0.931	0.553	-0.106	-0.891	0.666	-0.076	-0.799	0.655
Customary International law	0.219	-1.248	1.69	0.244	-1.22	1.729	0.675	-0.452	1.811	0.673	-0.424	1.791
Enlargement year 1973	-0.22	-1.722	1.256	-0.253	-1.746	1.224	0.012	-1.308	1.300	-0.016	-1.293	1.243
Enlargement year 1981	0.598	-0.604	1.594	0.482	-0.639	1.429	0.238	-0.848	1.280	0.166	-0.869	1.207
Enlargement year 1995	0.803	-0.521	1.983	0.741	-0.538	1.864	1.235	0.192	2.188	1.119	0.131	2.024
Enlargement year 2004							0.090	-1.357	1.539	0.004	-1.432	1.466
New member state							0.100	-1.352	1.553	-0.004	-1.476	1.447
σ_i		0.44			0.38			0.5			0.47	
WAIC		978.98			980.12			1133.59			1135.62	
$R^2_{glmm(m)}$		0.439			0.408			0.457			0.389	
$R^2_{glmm(c)}$		0.709			0.692			0.725			0.689	
Heidelberger and Welch Stationary test (σ_i)		p_{value} =			p_{value} =			p_{value} =			p_{value} =	
Heidelberger and Welch Halfwidth test (σ_i)		Mean= Halfwidth=			Mean= Halfwidth=			Mean= Halfwidth=			Mean= Halfwidth=	
Number of iterations	16000						16000					
Number of country levels	10						17					
Original sample size	169						215					
Weak Instrument test	F-statistic = 65.387***						F-statistic = 89.339***					
Hausman-Wu test	F-statistic = 15.082***						F-statistic = 21.205***					

Note: Priors have been specified (informative priors) based on the available information of the parameter and the response distribution corresponds to the negative-binomial distribution;
Constant includes the reference category *EU Founding Member States* (Belgium, Germany, France, Luxembourg, Italy and Netherlands).

C Descriptive Statistics: Figures

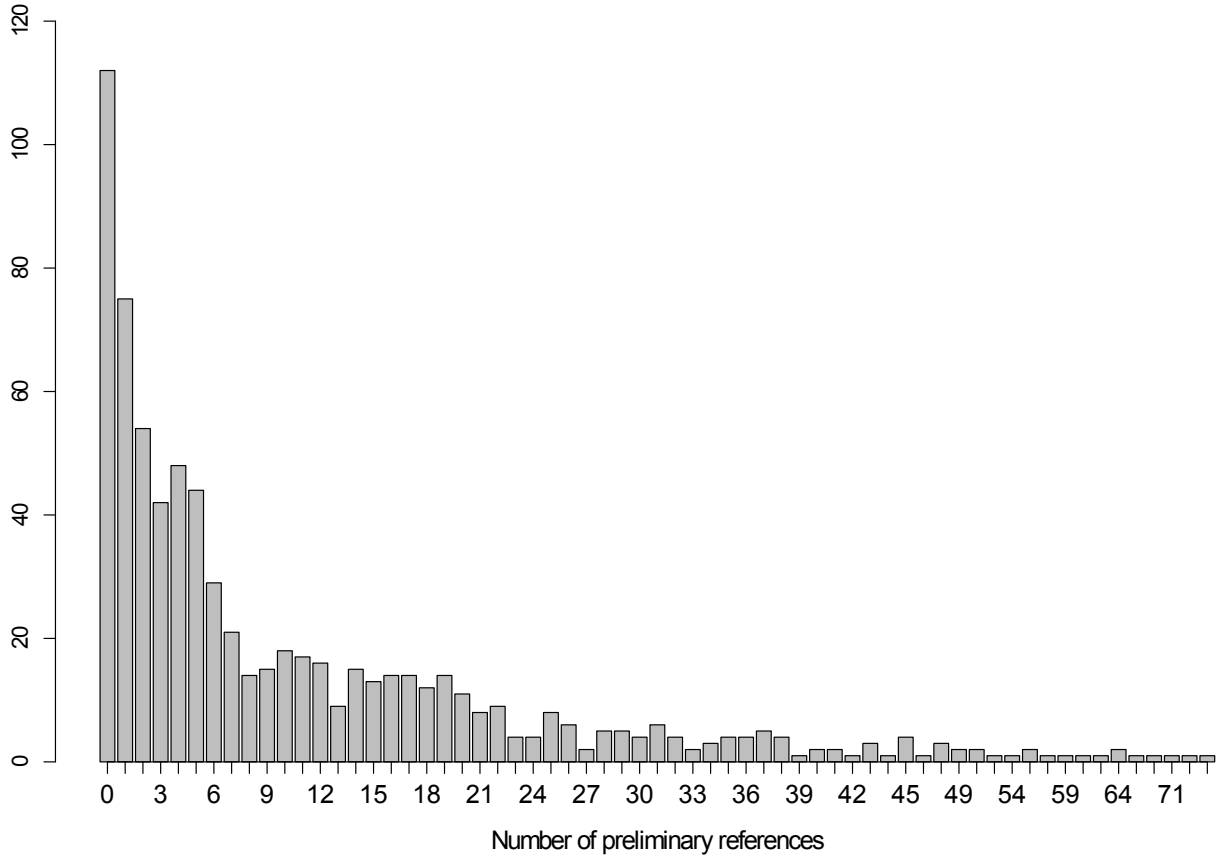
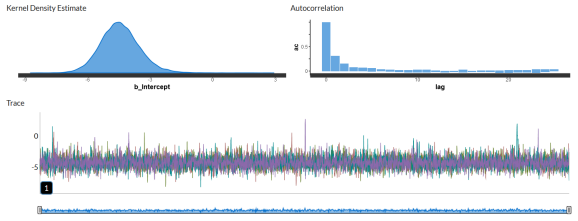


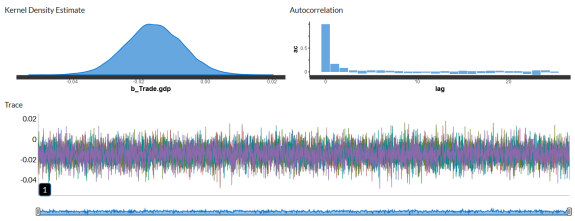
Figure 11: Frequency of number of preliminary references

D Diagnostics of Bayesian models

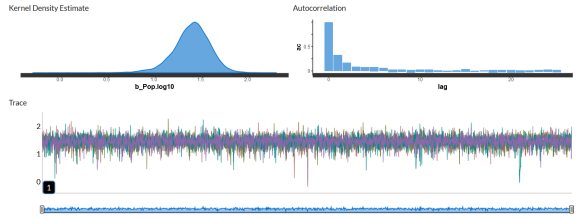
D.1 Model performance (*Ia*)



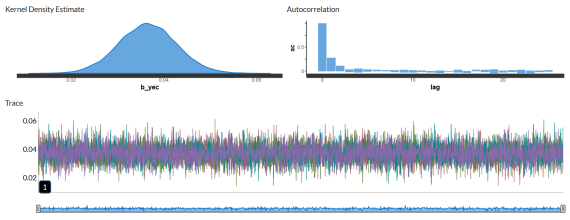
(a) Intercept



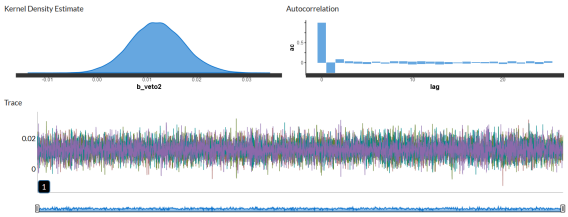
(b) Trade.gdp



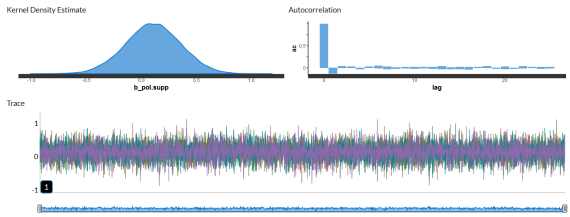
(c) Population



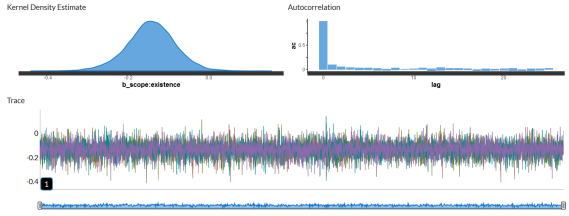
(d) Years of membership



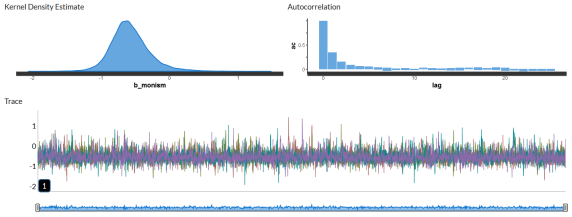
(e) Political Fragmentation (Veto)



(f) EU support

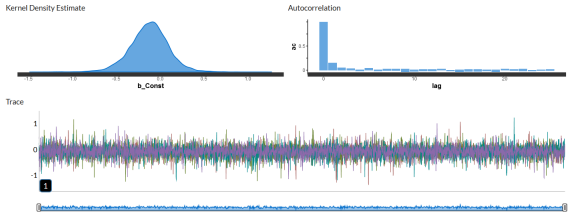


(g) Judicial review

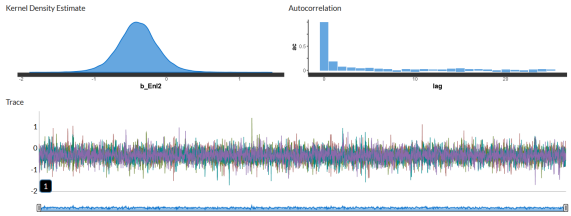


(h) Monism

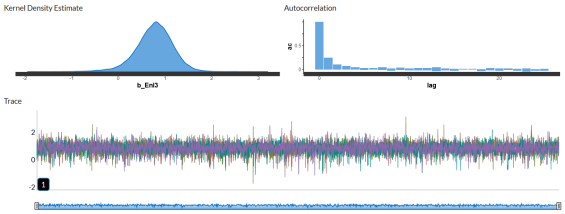
Figure 12: Model (Ia) diagnostics: Trace, density and autocorrelation plots of the first block of covariates



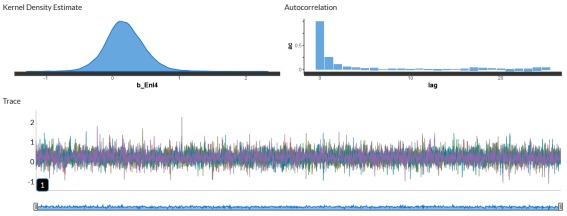
(a) Constitution



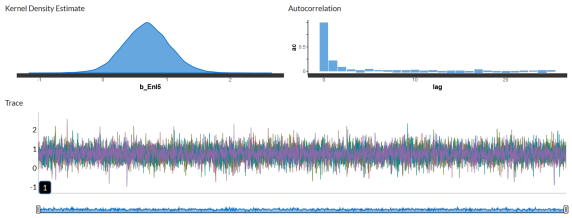
(b) Enlargement years 1973



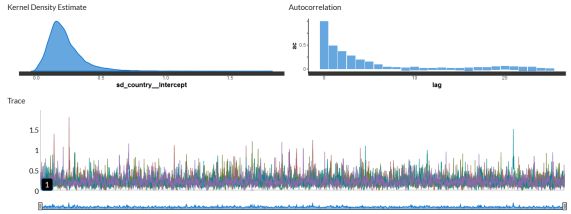
(c) Enlargement years 1981



(d) Enlargement years 1986



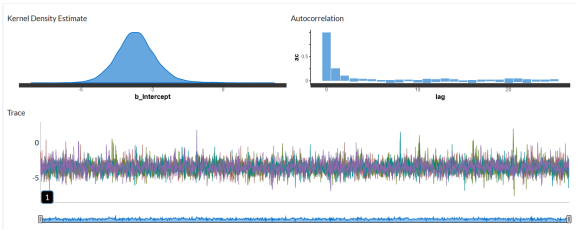
(e) Enlargement years 1995



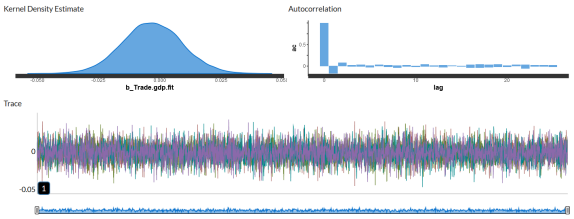
(f) Country

Figure 13: Model (Ia) diagnostics: Trace, density and autocorrelation plots of the second block of covariates

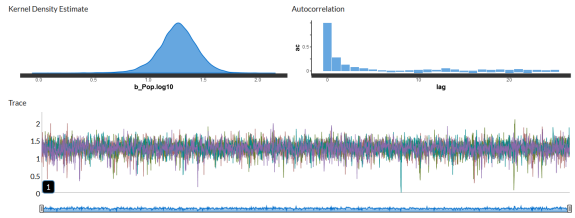
D.2 Model performance (*Ib*)



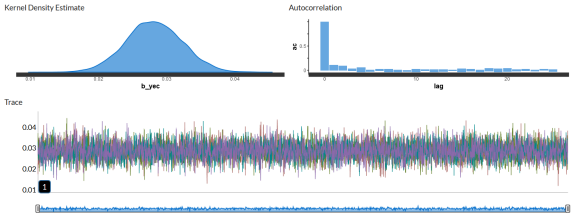
(a) Intercept



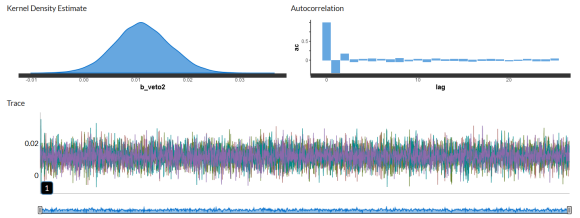
(b) Trade.gdp



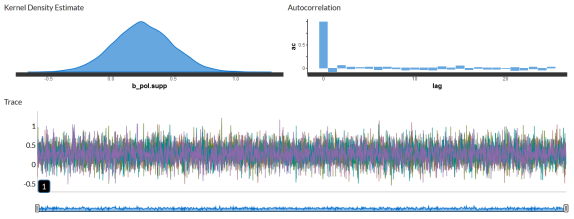
(c) Population



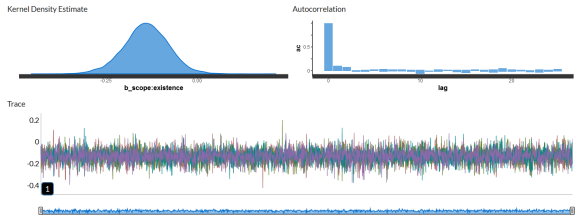
(d) Years of membership



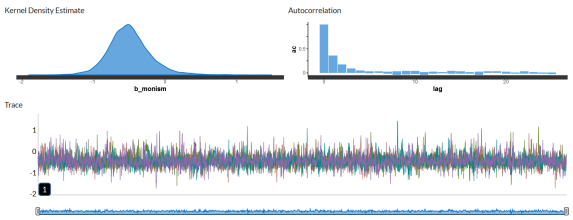
(e) Political Fragmentation (Veto)



(f) EU support

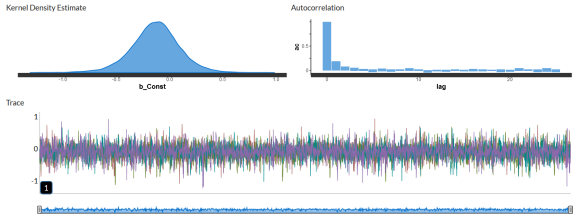


(g) Judicial review

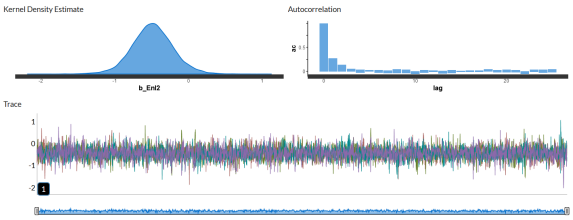


(h) Monism

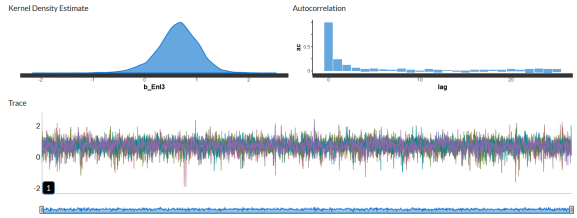
Figure 14: Model (Ib) diagnostics: Trace, density and autocorrelation plots of the first block of covariates



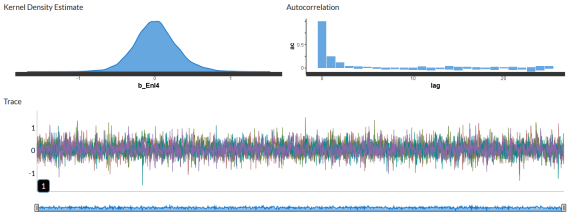
(a) Constitution



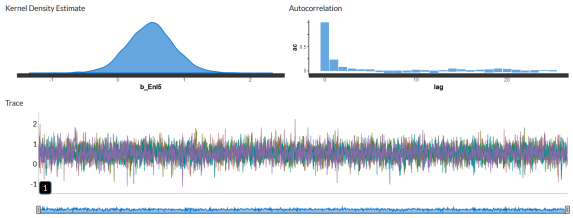
(b) Enlargement years 1973



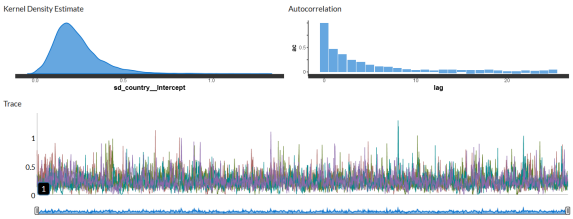
(c) Enlargement years 1981



(d) Enlargement years 1986



(e) Enlargement years 1995



(f) Country

Figure 15: Model (Ia) diagnostics: Trace, density and autocorrelation plots of the second block of covariates

E Acronyms

AT Austria

BG Bulgaria

CY Cyprus

CZ Czech Republic

DK Denmark

BE Belgium

DE Germany

DK Denmark

EE Estonia

ES Spain

EL Greece

FI Finland

FR France

HR Croatia

HU Hungary

IR Ireland

LI Lithuania

LV Latvia

LU Luxembourg

NL Netherlands

MT Malta

PL Poland

PT Portugal

RO Romania

SK Slovakia

SL Slovenia

UK United Kingdom

ECJ European Court of Justice