



The timely administration of justice: using computational simulations to evaluate institutional reforms at the CJEU

Joshua C. Fjelstul, Matthew Gabel & Clifford J. Carrubba

To cite this article: Joshua C. Fjelstul, Matthew Gabel & Clifford J. Carrubba (2022): The timely administration of justice: using computational simulations to evaluate institutional reforms at the CJEU, Journal of European Public Policy, DOI: [10.1080/13501763.2022.2113115](https://doi.org/10.1080/13501763.2022.2113115)

To link to this article: <https://doi.org/10.1080/13501763.2022.2113115>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



[View supplementary material](#)



Published online: 25 Aug 2022.



[Submit your article to this journal](#)



Article views: 640



[View related articles](#)



[View Crossmark data](#)

The timely administration of justice: using computational simulations to evaluate institutional reforms at the CJEU

Joshua C. Fjelstul ^{a,b}, Matthew Gabel^c and Clifford J. Carrubba^d

^aPolitical Science and International Relations, The University of Geneva, Geneva, Switzerland; ^bARENA Centre for European Studies, The University of Oslo, Oslo, Norway; ^cPolitical Science, Washington University in St. Louis, St. Louis, MO, USA; ^dPolitical Science, Emory University, Atlanta, GA, USA


ABSTRACT

The Court of Justice of the European Union (CJEU) faces a growing backlog of cases and delays in administering justice. Identifying the factors that prolong the resolution of cases is critical for designing successful institutional reforms. In this article, we use a Bayesian model to identify the characteristics of cases, judges, and the Court's internal procedures that affect the duration of cases. Then, we run computational simulations that use our model to estimate the likely effect of potential institutional reforms aimed at reducing the Court's backlog. We offer policy recommendations based on our findings. Our methodology can be tailored to study the efficiency of other domestic and international courts.

KEYWORDS European Union; Court of Justice of the European Union; administrative efficiency; institutional reforms; computational simulations

The primary function of the Court of Justice of the European Union (CJEU) is to rule on disputes in a timely manner. However, with the dramatic growth in cases over time, the CJEU has struggled to manage its docket, resulting in a substantial backlog of cases and delayed rulings. In 2018, the backlog at the Court of Justice (CJ) was well over 600 cases and the average CJ case took 17 months to decide. This is normatively undesirable, as 'justice delayed is justice denied.'¹ The CJEU has explicitly recognized this concern. President Due considered timely rulings as the first goal of delivering justice (European Court of Justice, 1995, 157). Former President Rodriguez worried that lengthy delays would dissuade national courts from referring cases to the CJEU and thereby undermine the uniform application of EU law (European Court of

CONTACT Joshua C. Fjelstul  joshua.fjelstul@unige.ch

 Supplemental data for this article can be accessed online at <http://dx.doi.org/10.1080/13501763.2022.2113115>.

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

Justice, 1995, 11). And, the EU panel that reviews judicial nominees for re-appointment to the CJEU explicitly considers the duration of cases and the productivity of the nominees (255 Comité, 2019, 6).

Not surprisingly, scholars, EU officials, and the CJEU itself have proposed reforms designed to expedite the review of cases and reduce the backlog (Dashwood & Angus, 2001; European Court of Auditors, 2017). Recognizing that the supply of cases, particularly through the preliminary reference procedure, is unlikely to abate, reform proposals have focused on expanding resources, dividing the work among the judges through the system of chambers, and procedural adjustments, such as reducing the use of Advocate General (AG) opinions. Surprisingly, few of the reform proposals are based on systematic empirical evidence about the sources of delay. Indeed, perhaps the best recent analysis consists of recommendations based on a 2017 review of a sample of 30 cases by the Court of Auditors (European Court of Auditors, 2017).

In this article, we use data on all CJ (1954–2018) cases (1) to identify the characteristics of cases, judges, and the internal procedures of the CJ that affect the duration of CJ cases, and (2) to estimate the likely effect of potential institutional reforms aimed at reducing case duration and the Court's backlog. First, we use a Bayesian model to estimate the correlation between these characteristics and the duration of CJ cases. Then, we run computational simulations based on our model to estimate the relative effects of potential institutional reforms on the average duration of cases and the size of the backlog.

Based on our empirical analysis, we identify four potential institutional reforms that we expect would improve the Court's productivity: (1) increasing the number of judges at the Court, (2) increasing the Court's use of small chambers, (3), decreasing the Court's use of AG opinions, and (4) increasing the Court's information technology (IT) expenditures, which go towards systems for managing cases. Our simulation provides an empirically-informed prediction about the extent to which each of these potential reforms would reduce the size of the backlog and the duration of cases.

We contribute to the emerging literature on the internal organization of the CJEU (e.g., Brekke et al., 2022; Cheruvu, 2019; Fjelstul, 2019; Frankenreiter, 2017, 2018; Hermansen, 2020; Malecki, 2012) by considering how the characteristics of cases, judges, and the Court's internal procedures affect its ability to process cases and publish judgments in a timely manner. We also contribute to the recent literature in political science and public administration about the effectiveness of (proposed) institutional reforms to increase the efficiency of legal processes (e.g., Cheruvu & Fjelstul, 2021).

The productivity of the court

The EU recognizes that delay in the processing of cases and the size of the backlog at the CJEU are serious logistical and political problems. It has

taken steps designed to address these interrelated problems and is considering further reforms to that end. For example, in 2017, the CJEU commissioned a report from the European Court of Auditors – an independent EU institution tasked with investigating and auditing other EU institutions' management of the EU's financial resources – to assess how the CJEU can improve the efficiency of the EU's judicial process by better utilizing the considerable resources (81 judges, 2,235 employees, and an annual budget of 437 million euros) it already has (Court of Justice of the European Union, 2021).

While we generally portray increasing the productivity of the Court as normatively good, we recognize that some means to that end may have important negative externalities. For example, hearing cases in smaller chambers to reduce the workload of judges could decrease the consistency of EU law (Fjellstul, 2022). There could also be a tradeoff between productivity and quality (Yeung et al., 2022).

Some delays might actually be valuable. The Court could delay cases that present complex, novel legal questions while it develops precedent. In addition, the Court could prioritize cases because of their political salience (e.g., case C-370/12 on the legality of the European Stability Mechanism, which was heard by 27 judges and decided in less than four months), or it could delay cases to avoid the appearance of interfering in politics. In our empirical analysis, we consider how legal and political complexity affect duration, and we develop our policy recommendations with this fuller set of considerations in mind.

Concerns about the productivity of the CJ center around two indicators: the average duration of cases and the size of the backlog. These are separate but closely related metrics. The average duration of cases is the average period of time (we measure duration in months) between the date a case is lodged and the date the Court publishes a judgment. The CJ's backlog is the set of all cases that are open (i.e., have been lodged but not decided) at a given point in time.

We emphasize that a large backlog does not necessarily equate to delay in the processing of cases. The size of the backlog at any point in time depends on the number of incoming cases and the duration of those cases. It is possible for the backlog to be growing even as the duration of cases is decreasing, and *vice versa*. But clearly, lowering case duration has a salutary effect on the backlog.

In this article, we are ultimately interested in the duration of cases, as this is what matters to litigants, but the size of the backlog plays an important role in our theoretical account of case duration. We expect that the size of the backlog actually affects the duration of cases. In particular, we expect there to be a positive feedback loop: as cases take longer, the backlog grows, which increases the workload of judges, which further increases how long cases take to decide. As the workload of judges increases, judges have to

simultaneously handle more cases. We expect this multitasking to slow down their processing of individual cases, and we provide empirical evidence of this in our analysis.

Figure 1 plots the average duration of CJ cases (Panel A) and the size of the CJ's backlog (Panel B) by month from the founding of the Court in 1952 through 2018. The average duration of cases increased substantially between the mid 1970s and the late 1990s. In the early 2000s, this trend reversed, and since the mid 2000s, it has remained steady at approximately 18 months. The backlog has increased substantially since the mid 1990s as the EU has enlarged and as the number of incoming cases has increased. Currently, the backlog is well over 600 cases.

Figure 2 shows the composition of the backlog by legal procedure and by the size of the chamber that hears the case, which is a proxy for the relative importance of the case. The backlog comprises all open cases, so its composition is primarily determined by the distribution of incoming cases. The CJ does not have docket control—it must decide all admissible cases that are lodged—and therefore has little control over the composition of the backlog. Since the early 2000s, the backlog has increasingly become composed of references for a preliminary ruling heard by 5-judge chambers—substantively important cases that play an important role in the development of EU law.

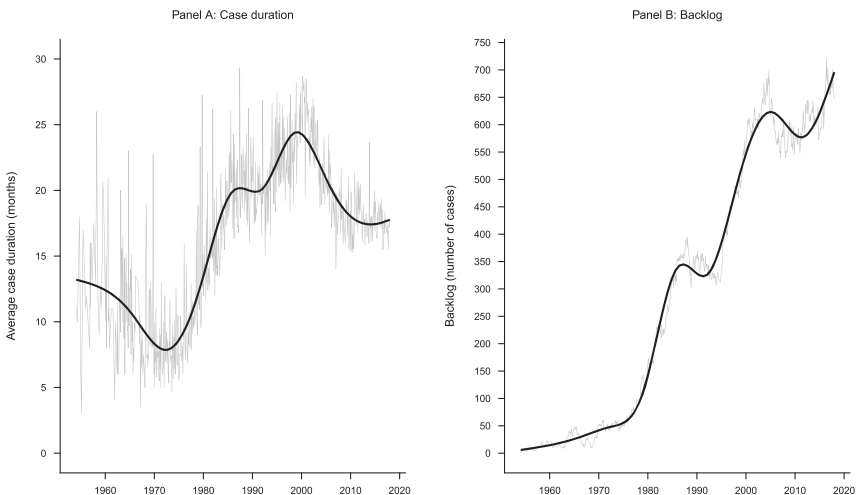


Figure 1. This figure shows the average duration of CJ cases (Panel A) and the backlog of CJ cases (Panel B) at the CJ by month from 1954 to 2018. We include a generalized additive model (GAM) smoother.

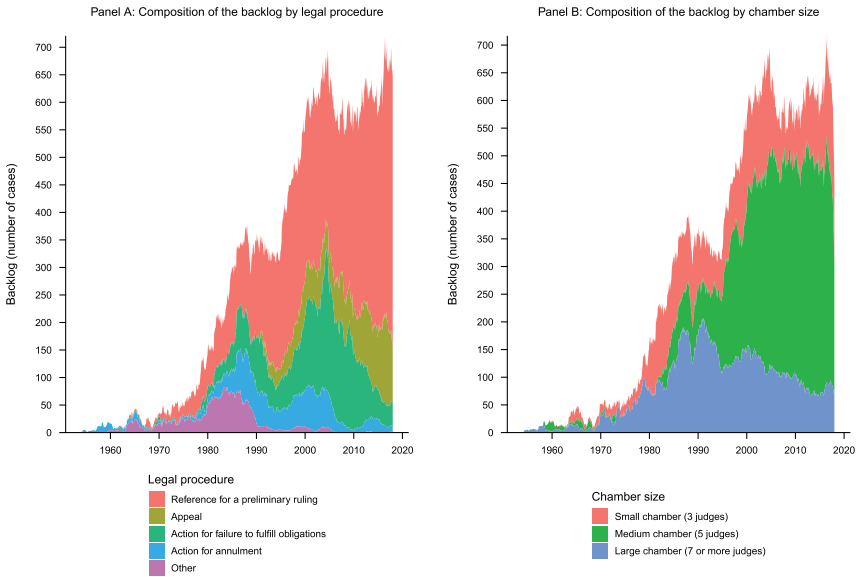


Figure 2. This figure shows the composition of the backlog by legal procedure and by chamber size, which captures the importance of the case.

Potential procedural sources of delay

In this section, we review the Court's Rules of Procedure to both describe the timeline of case review and identify institutional features that could influence case duration and the backlog. See Lenaerts et al. (2014) and Brekke et al. (2022) for a more detailed overview of the Court's procedures. The Court hears references for a preliminary ruling, where a national court refers questions about EU law to the Court, and direct actions, which are filed by applicants.² As soon as a case is filed, the President of the Court assigns a Judge-Rapporteur, who is responsible for managing the case and writing the draft judgment. Because of their central role in the managing of cases, the duration of a case depends on how effectively the Judge-Rapporteur can process their workload. An Advocate General (AG) is also assigned at this time. AGs are judge-like legal experts who provide opinions to the Court in many cases.

When a case is filed, the parties submit a variety of procedural documents to the Court. Third-party actors can also submit written observations in references for a preliminary ruling and can intervene in support of one of the parties in direct actions. This stage of the procedure generates an enormous amount of paperwork. In 2020, over 159,000 procedural documents were filed, requiring the Court's 600 lawyer-linguists to translate over 1.1 millions of pages of text (Court of Justice of the European Union, 2021).

Translation requirements can slow the legal proceedings in several ways. First, the initial pleadings are translated into working languages of the court, and in the case of preliminary references the national court filing is translated into all official EU languages and distributed to the member-state governments. Second, third-party interventions, observations, and supporting documents may require translation. Third, cases can be prolonged due to delays in translation by the EU institutions (e.g., the Commission) that regularly participate in cases (European Court of Auditors, 2017, 27). Not surprisingly, translation issues have often been cited as a major source of delay (Dashwood & Angus, 2001, 120; European Court of Auditors, 2017, 27; Strasser, 1995, 63).

The Judge-Rapporteur reviews all of the documents that are submitted and presents a preliminary report on the case at a general meeting of the Court. The Court decides what formation will hear the case (i.e., the number of judges that will be on the panel). Prior to 2001, the AG assigned to the case then wrote an opinion on the case, which would be presented to the Court before it decided the case. Since 2001, the Court has had discretion about whether to request an AG opinion.

The Judge-Rapporteur must issue a preliminary report on each case in a timely manner, regardless of the backlog of cases on their desk. Thus, as more cases come in, the Judge-Rapporteur must spread their time ever more thinly across the cases they are responsible for. The size of the backlog also causes delays in AG opinions. In fact, in a review of 30 cases by the Court of Auditors, the workload of the Judge-Rapporteur and AG was one of the strongest sources of delay (European Court of Auditors, 2017, 31).

After the oral hearing, the judges deliberate, vote (using a simple majority decision rule), and publish a decision. The deliberations are held in French, which is the working language of the Court. Cheruvu (2019), in the only study that has specifically looked at the duration of CJ cases, finds that cases managed by Francophone Judge-Rapporteurs (i.e., judges from France, Luxembourg, and French-speaking Belgium) are shorter than cases managed by Judge-Rapporteurs who are not native French-speakers. Finally, the Court's judgment is written by the Judge-Rapporteur.

In sum, the judicial procedure involves several steps where institutional features can potentially cause delay. These include the workload of judges and AGs, the Court's investment in case management systems to help Judge-Rapporteurs, the size of the panel hearing the case, the translation requirements, and whether there is an AG opinion. We examine how these institutional features, as well as case- and judge-specific attributes, relate to case duration and backlog in the next section.

A Bayesian model of case duration

We begin our analysis by developing a Bayesian model to estimate the correlation between various characteristics of cases, judges, and features of the Court's internal procedures and the duration of cases. Drawing on the recent literature and our own analysis of the Court's Rules of Procedure, we operationalize measures of the variables that we expect to be associated with the duration of CJ cases. Our analysis has two goals: (1) identify variables that are associated with case duration, and (2) develop a model that accurately predicts case duration. We will use this model to predict the duration of cases in our simulations to evaluate the effect of potential institutional reforms. While our model is not specified to draw inferences about the causal effects of particular independent variables, we do identify a variety of variables that are strongly correlated with CJ case duration.

Our model includes 23 variables that we expect to affect the duration of CJ cases. We group these variables into 4 categories: (1) variables about the internal procedure of the CJ, (2) variables about the content of cases, (3) variables about the personnel involved in the case, (4) and variables about the legal procedures in the case. Our sample includes the universe of CJ cases that lead to judgments before December 31, 2018. Our case-level variables are based on metadata from EUR-Lex.

Measurement

First, we include 8 variables about the Court's internal procedure. The two most important variables in the model are *Judge workload* and *AG workload*. We measure the workload of judges as the number of open cases (i.e., the size of the Court's backlog) divided by the number of judges. Note that *Judge workload* also captures the workload of each judge's référendaires, who assist judges with their work (Kenney, 2000).³ Similarly, we measure the workload of AGs as the number of open cases in which there is an AG opinion divided by the number of AGs (as the workload of the AG with respect to a case is much higher when there is an AG opinion). We expect the Court to operate more efficiently when the workload of judges and AGs is smaller, as there is less multitasking.

These two workload variables will play a critical role in our simulation because they create a feedback loop: reductions in the duration of cases caused by other variables will decrease the backlog, which will decrease the workload of judges and AGs, which will decrease the duration of future cases. This allows our simulation to capture the long-term effects of potential reforms, as their effects build over time.

Panel size is the number of judges on the panel that hear the case. We expect cases heard by a larger panel to take longer to decide for several

reasons. In terms of logistics, there are more schedules to coordinate and debate among the judges could take longer. But panel size also measures the political saliency of the case. The Court tends to reserve larger panels for politically important cases, which may take longer (Kelemen, 2012).

AG opinion measures whether there was an AG opinion in the case, which is an extra procedural step that could cause delay. Since 2001, the Court has had the discretion to request an AG opinion, and it has increasingly opted against an opinion. Currently, there is an AG opinion in approximately 60 percent of references for a preliminary ruling and 30 percent of direct actions (Brekke et al., 2022).

IT expenditures is the Court's total annual expenditures in millions of euros on IT systems. The Court's IT budget goes towards systems for managing cases (European Court of Auditors, 2017). Therefore, we expect an increase in IT expenditures to improve the Court's ability to process cases efficiently. We collect data on IT expenditures from the EU's annual budgets (available since 1961), which are available on EUR-Lex. We do not include a separate measure of the number of staff because the number of staff is highly correlated with the budget ($r > 0.9$).

The Court frequently joins similar cases together, and *Joined cases* is the number of cases that have been joined (coded 1 for stand-alone cases). We expect that joined cases will take longer to process because judges have to deal with more material during the written stage and have to analyze case facts for more litigants.

Authentic languages measures the number of authentic languages in a case, which is the number of official EU languages spoken by the litigants. We expect cases with a higher number of authentic languages to take longer because procedural documents will need to be translated into more languages (European Court of Auditors, 2017).

Finally, we control for the creation of the General Court (then called the Court of First Instance) in 1989 with a dummy variable that equals 1 for all cases after 1989. The creation of the General Court was intended to reduce the workload of the Court of Justice, but it also created more demand for shared resources, like translation services.

Second, we include 6 case-level variables about the content of the case. On balance, we expect that the Court will take longer to decide complex cases than straightforward cases. This is exactly what the Court of Auditors found in their study of 30 cases (European Court of Auditors, 2017, 31). We develop variables to measure several dimensions of complexity. See the Supporting Information for details.

Citations is the number of case law citations in the Court's judgment. Judgments that cite more case law will, on balance, require more research by the référendaires. Relatedly, *Sources of law* is the number of sources of law (i.e., treaty law, legislation, and case law) represented across the judgment's

citations. Cases that deal with multiple sources of law require a higher degree of synthesis, which takes time.

Policy areas is the number of policy areas that the case deals with. When a case deals with multiple policy areas, it requires judges to develop more substantive expertise and to synthesize more areas of law. *Legal procedures* is the number of legal procedures associated with a case. Cases with multiple legal procedures are rare, but we expect them to take longer because the judges will have to decide multiple legal questions.

Principles of law indicates whether the case deals with general principles of law. We expect these cases to be less complex because the law is well-established and more likely to be consistent across member states' legal traditions. Judges are more likely to be applying existing legal rules than developing new ones. *Legal order* indicates whether the case deals with the relationship between EU law and member state law. We expect these cases to take longer to decide because they generally deal with conflicts between EU law and member state law that are politically sensitive.

Third, we include 3 variables about the personnel assigned to the case. *JR experience* is the number of years that the Judge-Rapporteur has served on the Court when the case is opened. *AG experience* is the same, but for AGs. We expect more experienced judges and AGs to have developed more effective strategies for managing their workloads. They are more familiar with the Court's internal procedure and are more practiced at writing judgments (Strasser, 1995, 60). *Francophone* indicates whether the Judge-Rapporteur is from a French-speaking member state (France, Luxembourg, or Belgium). Cheruvu (2019) shows that Francophone judges decide cases faster than non-Francophone judges because they have native proficiency in the Court's working language.

Fourth, we include 6 variables that control for the legal procedure. There are significant differences in case duration across legal procedures (Strasser, 1995). Due to shorter deadlines and their general prioritization by the Court, references for a preliminary ruling tend to be shorter than direct actions, for example (Strasser, 1995, 65). See the Supporting Information for a review of the legal procedures.

Analysis

Our dependent variable is the duration of a case in months. We estimate a linear Bayesian regression.⁴ Using a Bayesian model is useful because it allows us to incorporate uncertainty about the marginal effect of each independent variable into our simulations, allowing us to plot credible intervals for our predictions of the size of the backlog and the average duration of cases in counterfactual scenarios. Figure 3 shows a coefficient plot that summarizes the results of our model (including 90 percent credible intervals).

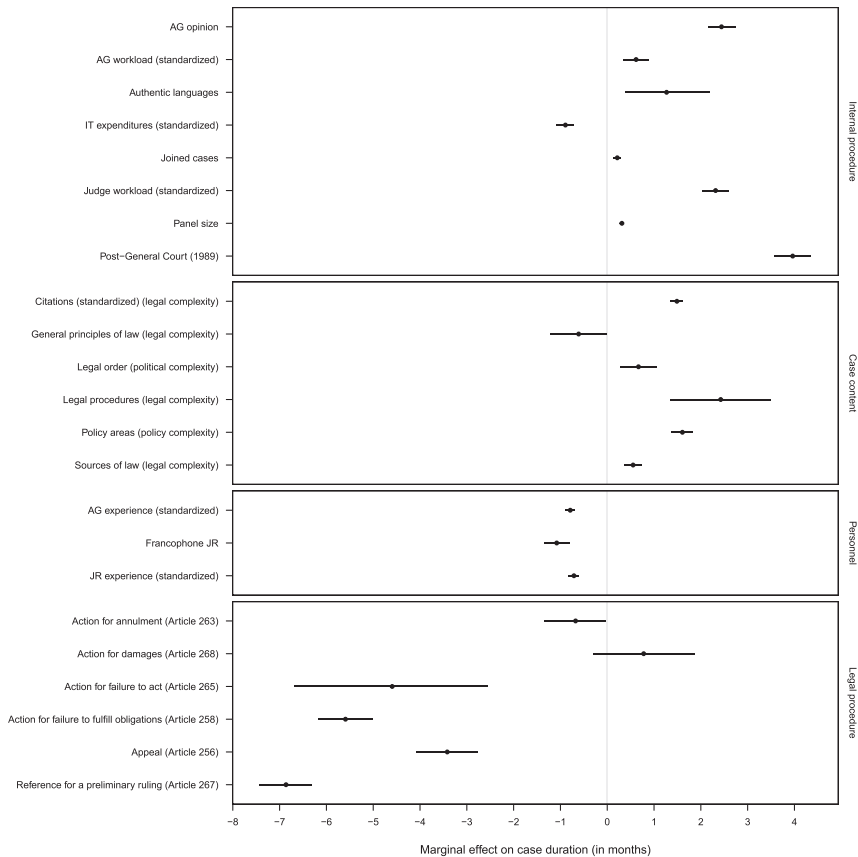


Figure 3. This figure shows the estimated marginal effect for each variable in our Bayesian regression model of case duration (with 90 percent credible intervals).

To assess model fit, we calculate a Bayesian version of the R^2 statistic proposed by Gelman et al. (2019). The Bayesian R^2 has a median of 0.372 and a standard deviation of 0.006. In our analysis of our computational simulation, which is based on this model, we will show that the predictive power of the model is sufficient to accurately predict the Court's backlog over time.

Looking at the first panel of Figure 3, on internal procedures, we find that the average workload of judges and AGs matters. A one standard deviation increase in *Judge workload* (8 cases) is associated with a 2.32 month increase in the duration of a case. This is equivalent an increase of 0.28 months per case. The effect is smaller for AGs. A one standard deviation increase in *AG workload* (10 cases) is associated with a 0.62 month increase (an increase of 0.06 months per case).

The marginal effect of *Panel size* is modest. A two-judge increase in the size of the panel (panels have to be an odd number of judges, so the minimum increase is two) is associated with a 0.62 month increase in case duration.

In contract, the presence of an AG opinion has a substantial effect and is associated with a 2.44 month increase in case duration. A one standard deviation increase in *IT expenditures* (6 million euros) is associated with a 0.89 month decrease in case duration. A one million euro increase leads to a 0.15 month decrease in case duration.

We find that cases after the creation of the General Court take 3.96 months longer than those before. This is a puzzling finding, as we might expect the GC to ease the burden on the CJ, making it more productive (e.g., Yeung et al., 2022). One explanation could be that the GC increased the burden on shared administrative resources, which did not expand quickly enough to serve the needs of both courts. However, *General Court* is only a dummy, and not a causal estimate, so we have to interpret this cautiously. Future research should try to identify the causal effect of the creation of the General Court.

Looking at the second panel of Figure 3, we find that that case complexity is strongly associated with case duration. A one standard deviation increase in the number of citations (approximately 7 citations) is associated with a 1.49 month increase in duration (an increase of 0.2 months per citation), an additional policy area is associated with a 1.61 month increase, an additional source of law is associated with a 0.55 month increase, and an additional legal procedure is associated with a 2.43 month increase. Cases that deal with the EU legal order are 0.67 months longer, on average, and cases that deal with general principles of law are 0.61 months shorter.

Looking at the third panel of Figure 3, we find that the attributes of judges and AGs also matter. A one standard deviation increase in *JR experience* (3.89 years at the CJ) is associated with a 0.71 month decrease in duration (a decrease of 0.18 months per year), while a one standard deviation increase in *AG experience* (4.33 years at the CJ) is associated with a 0.79 month decrease (also a decrease of 0.18 months per year). Cases with a Francophone Judge-Rapporteur are 1.08 months shorter on average.

Our estimates do not have a causal interpretation, but we have identified a variety of variables that are strongly correlated with case duration. We find that complex cases take longer, that more experienced judges and AGs work faster, and that cases take longer to decide when the workload of judges is higher. We also find that using a larger panel and having an AG opinion, both of which are discretionary, increase case duration. Finally, we find that higher IT expenditures, which fund case management systems, are associated with shorter cases.

In sum, we find that a substantial source of delay involves factors beyond case and judge characteristics. This is important, as it means that case duration could potentially be reduced through institutional reforms without relying on changes in how the court deals with politically salient and legally complex cases. In the next section we use simulations to evaluate several such institutional reforms.

Institutional reforms

Building on our findings in the previous section, we focus our analysis on four potential reforms that we expect would decrease average case duration and the size of the backlog: (1) increasing the size of the court, (2) increasing the Court's use of small chambers, (3), decreasing the Court's use of AG opinions, and (4) increasing the Court's IT expenditures. We also discuss the mechanisms by which each reform would operate.

First, we simulate an increase in the number of judges (Reform 1a) and AGs (Reform 1b) on the Court. Currently, there is one judge per member state, so we simulate an increase to two judges per member state. Since 2016, the CJEU has gradually increased the number of judges at the GC from one per member state to two per member state, so doing the same at the CJ is an obvious reform to consider. The GC reform also demonstrates that doubling the number of judges is politically feasible. Separately, we simulate doubling the number of AGs, which has increased over time. Originally there were 2, and then the number increased to 4 in 1968, to 8 in 1978, and to 11 in 1992.

Second, we simulate the Court only using 3-judge panels to decide cases (Reform 2). Since the early 2000s, the Court has increasingly relied on 3-judge and 5-judge chambers to reduce the workload of the judges, so this is just a more extreme version of what the Court is already doing. This counterfactual establishes an upper-bound on the effect of reforms to the chamber system on average case duration and the size of the backlog. We also compare this counterfactual with a counterfactual in which the Court uses a 15-judge chamber (the current size of the Grand Chamber) to hear every case (Counterfactual 1). This gives us an estimate of how large the backlog could have been had the Court not moved away from larger formations.

Third, we simulate the Court eliminating AG opinions in all cases (Reform 3). Over the last decade, the Court has deliberately reduced the use of AG opinions to speed up the judicial process. Currently, the Court tends to reserve AG opinions for cases that raise novel legal issues. Reducing the number of AG opinions removes a time-consuming step of the Court's internal procedure, but it also reduces the average workload of AGs.

We assume that the Court does not replace AGs with another mechanism. We think this is a plausible assumption. AGs opinions are unique to the CJ – the GC does not have them, and judges at other constitutional courts decide cases without the help of an independent expert. Our model shows that, on balance, not having an AG opinion speeds up cases. However, as the Court increasingly reserves AG opinions for cases that present novel legal issues, it could be that the marginal effect of not having an AG opinion attenuates. In the Supporting Information, we investigate this possibility and find no evidence of this.

Fourth, we simulate an increase in IT expenditures, which fund systems for managing cases.⁵ We simulate a 50% increase (Reform 4a) and a 100% increase (Reform 4b) in IT expenditures. These are plausible increases. The Court's IT expenditures in 2020 were 23.2 million euros, which is a 93.7% increase from 2010 expenditures (12.0 million euros). Total expenditures in 2020 were 440.6 million euros, so IT expenditures only account for 5.3% of the total.⁶ Since 2018, total expenditures have increased by 30.6 million euros, which is a larger increase than a 100% increase in IT expenditures.⁷

We also simulate a reform package (Reform 5) that combines elements of each of these. The package includes: (1) an increase in the number of judges from one per member state to two per member state, (2) using 3-judge chambers to decide all cases, (3) eliminating AG opinions in all cases, and (4) a 50% increase in IT expenditures. This simulation captures an upper-bound on how much these proposed reforms could reduce average case duration and the size of the backlog, which will give us a better sense of how significant the reforms would need to be in order to get the backlog under control.

Computational simulations

We use computational simulations based on our Bayesian model to estimate the effect of the potential reforms we have identified above. We can think of our simulations as quasi-experiments. Each reform is a treatment. We choose a historical date on which to simulate the implementation of each reform. The unit of time in our simulations is a month, so the months before and after the reform goes into effect are the pre-treatment and post-treatment periods, respectively. We run a treatment simulation, analogous to a treatment group, in which we simulate the reform being implemented, and a control simulation, analogous to a control group, in which we simulate no reform. Both the treatment and control simulations are stochastic, so we run 100 iterations for each reform.

The outcome variables are the average duration of cases and the size of the backlog. The size of the backlog endogenously influences case duration via the workload of judges and AGs. We are interested in the average treatment effect on the treated (ATET), which is the average difference in average case duration and the size the backlog between each treatment simulation and the control simulation in the post-treatment period. This is the expected effect of each reform. We calculate the ATET for each month in the post-treatment period. The ATET is an average in the sense that it is the average effect across the 100 iterations. We plot the ATET over time across the post-treatment period, as the effect of a reform can build over time.

To run our quasi-experiments, we need to choose a historical date on which to simulate the implementation of each reform. This choice is arbitrary.

We choose May 2004, which is the date of the EU's enlargement into central and eastern Europe. This is a clear inflection point in the Court's history.

In each simulation, we loop through all of the empirical cases in chronological order and use our trained Bayesian model to predict the duration of each case. For cases in the pre-treatment period, we use the empirical case duration as our prediction. As we do this, we update the values of the workload variables (*Judge workload* and *AG workload*), which depend on the current backlog, based on the output of the simulation up to that point. This makes the simulation dynamic and allows reductions in the duration of individual cases to contribute towards a gradual reduction of the backlog, which reduces the workload of judges and AGs, and further reduces the duration of individual cases.

We draw predictions from the posterior predictive distribution instead of using the point predictions of the linear predictor, making our predictions stochastic. These predictions incorporate the residual error in the model and therefore have a higher variance than the predictions of the linear predictor. Using the posterior predictive distribution allows our simulation of the Court's backlog to incorporate uncertainty in our Bayesian model about the marginal effect of each variable.

In the control simulation, we simply use the empirical data for each case to predict case duration. In the various treatment simulations, we adjust the empirical data to capture the counterfactual we are interested in. For example, to simulate an increase in the size of the court from one judge per member state to two judges per member state (Reform 1), we double the number of judges for all cases in the post-treatment period, recalculate the *Judge workload* variable based on that higher number of judges, and use that adjusted data to predict case duration.

After we have predicted the duration of each case, we calculate average case duration and the size of the backlog. This yields two time series of monthly predictions—one for average case duration and one for the size of the backlog—for each iteration of the simulation. We calculate the average of these predictions by month across all of the iterations. We also calculate 90 percent credible intervals, which capture the uncertainty in our predictions. Since we use the predictive posterior distribution to stochastically predict case durations, and those predictions incorporate the residual error of the regression model, these credible intervals are conservative.

To validate our methodology, we compare the predictions of our control simulation (but using our model to predict case duration for all cases, both those in the pre-treatment period and the post-treatment period) with the empirical average case duration and the empirical size of the backlog. Figure 4 shows that the simulated and empirical values track together well. In terms of predicting the backlog, our simulation significantly outperforms a naive estimator that uses the median duration as the prediction for each

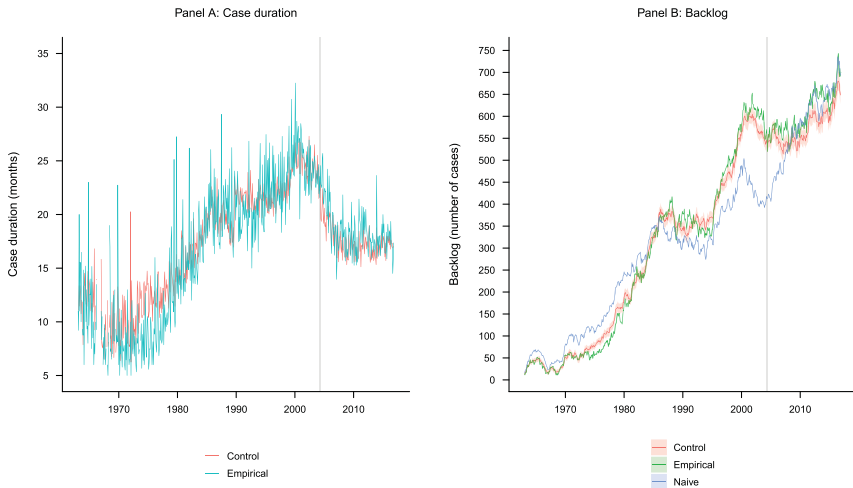


Figure 4. Panel A compares empirical average case duration to predicted average case duration in the control simulation. Panel B compares the empirical backlog to the predicted backlog.

case. This indicates that our model has enough predictive power that we can accurately reconstruct the empirical backlog using the model's predicted case durations. This increases our confidence that using our model to predict case durations under counterfactual scenarios will yield plausible estimates of average case duration and the size of the backlog.

Findings

In this section, we present the findings of our simulations. Figure 5 plots the predicted backlog for Reforms 1 to 4. This figure shows the period 1994–2018. We simulate the treatment as being applied in 2004 (indicated by the vertical line). We show 10 years of the pre-treatment period (1994–2002) so we can evaluate the degree to which each reform would have reduced average case duration and the size of the backlog relative to their pre-treatment values.

Figure 6 plots the estimated ATET for each reform by month, which is the vertical difference between the control simulation and the treatment simulations in Figure 5. Since this figure shows the estimated ATET, it shows only the post-treatment period (2004–2018). This figure allows us to compare the extent to which each reform would have reduced average case duration and the size of the backlog and how those effects change over time.

All of the simulated reforms reduce average case duration and the backlog. However, their substantive effects vary dramatically. Looking at Figure 6, the most effective reform is eliminating AG opinions. By the end of the post-

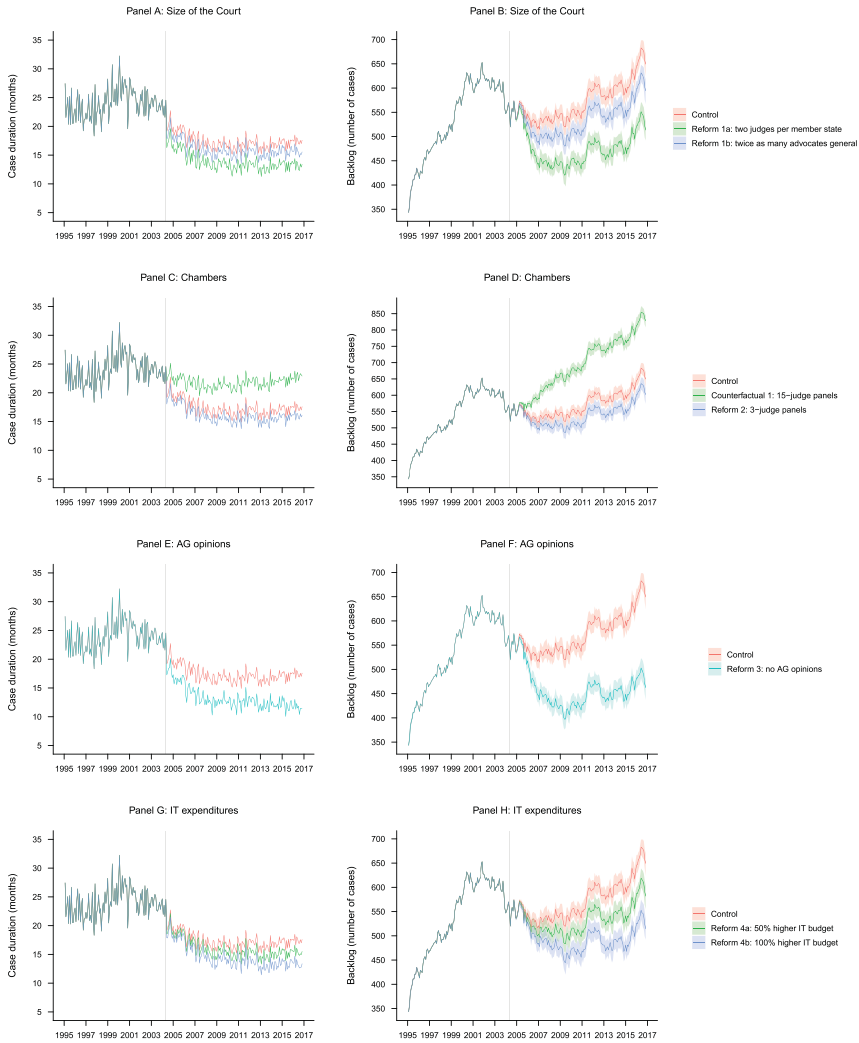


Figure 5. This figure shows the simulated effect of each reform on the average duration of cases and on the backlog.

treatment period, the ATET for average case duration is a decrease of around 6 months, and the ATET for the backlog is a decrease of 175 cases. This reform also makes an immediate difference. The ATET for the backlog after 5 years is a decrease of around 125 cases, which is the highest of any of the reforms. To the extent that CJ judges rely on research done by AGs, one consequence of reducing AG opinions is that it could transfer work to judges and their référendaires.

The second most effective reform is increasing the number of judges to two per member states (Reform 1a). By the end of the post-treatment

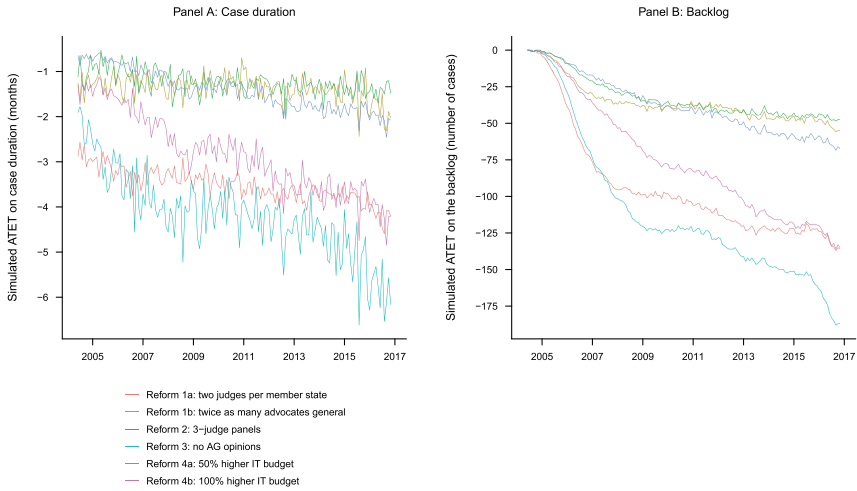


Figure 6. This figure shows the simulated ATET for the average duration of cases and the backlog by month across the post-treatment period for each reform.

period, the ATET for average case duration is a decrease of around 4 months and the ATET for the backlog is a decrease of approximately 140 cases. Doubling the number of judges has a substantially larger effect than doubling the number of AGs, indicating the the workload of judges is more of a bottleneck than the workload of AGs. One aspect of this reform that our methodology does not capture is that newly appointed judges would have no CJEU experience, which could undermine the efficiency gains. There could also be budget implications. If member states do not increase the Court’s budget to cover the salaries of new judges and their staffs, offsetting cuts could cause negative externalities in other areas.

The third most effective reform is increasing IT expenditures by 100 percent (Reform 4b). This reform takes longer to make a difference, but by the end of the post-treatment period, the ATETs are approximately the same as the ATETs for doubling the number of judges. Note that we assume constant marginal returns to IT spending, but there could be diminishing marginal returns that our model does not capture. In the Supporting Information, we estimate models that allow the effect of IT spending on case duration to be non-linear, but we do not find any evidence of diminishing marginal returns. It may be that they do not set in until expenditures rise higher than those that we observe empirically. Nevertheless, we need to be cautious in interpreting this result.

The least effective reforms are doubling the number of AGs (Reform 1b), mentioned above, and using only 3-judge chambers (Reform 2). Looking at

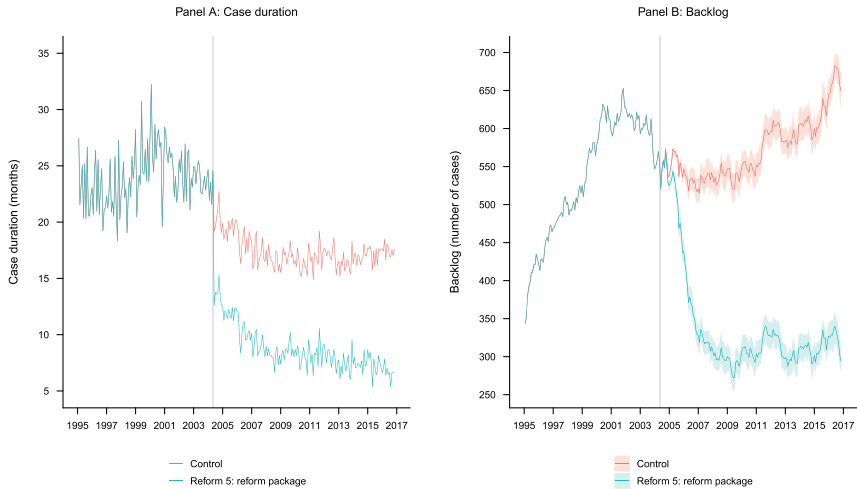


Figure 7. This figure shows the simulated effect of the reform package on the average duration of cases and on the backlog.

Panels C and D of Figure 5, using 3-judge chambers only leads to a modest improvement in average case duration and the size of the backlog. By the end of the post-treatment period, the ATET for the backlog is a reduction of approximately 50 cases. This suggests that the Court is already close to maximizing its use of chambers. Had the Court heard cases in 15-judges chambers, the backlog would have been far higher. As chamber size decreases, there is greater potential for inconsistency across similar cases (Fjelstul, 2022), so a negative side effect of this reform could be higher inconsistency in the application of EU law.

One important pattern to notice in Figure 5 is that, with all of the reforms, the backlog is still increasing over time, which means none of these reforms actually reduce the backlog to a sustainable level. By the end of the post-treatment period, the ATETs never get large enough for any of the reforms in Figure 5 to stop the backlog from trending upward. Looking at the reform package (see Figure 7), however, average case duration is trending downward and the backlog is stable over time. This implies that the Court actually has excess capacity under this set of reforms.

Conclusion

Our analysis indicates that case- and judge-specific factors are important contributors to the duration of cases and the backlog at the CJ. Given the realities of the appointment process and the fact that the Court does not have docket control, these factors are largely impervious to reform. But we also find that features of the Court's internal procedure play a substantial role in case

duration and are potentially strong candidates for successful reform. Specifically, our simulations of counterfactual procedural arrangements show that increasing the number of judges, eliminating AG opinions, and increasing funding for case management systems would have significantly reduced average case duration and the size of the backlog.

In considering specific policy reforms, we recognize that procedural changes that are likely to reduce delay may have adverse effects. For example, the CJEU also would like to produce high quality and consistent judgements. We also recognize that some delays may reflect the political saliency or the legal complexity of the case, and we would not want to reform procedures such that they restrict the Court's discretion to extend their consideration of the cases for those reasons.

With that in mind, our findings point to two policy recommendations. First, the EU should increase the number of judges at the CJ from one per member state to two per member state, as it has already done for the GC. Second, the Court should invest in better case management systems. We think these reforms, which are about increasing resources, are less likely than reforms that involve procedural changes to impact the quality and consistency of judicial decision-making. However, policy-makers should carefully consider potential side effects that our methodology does not directly capture.

Our simulation predicts that eliminating AG opinions would have had the largest effect of any potential reform we consider. However, AG opinions play an important role in the judicial process, and eliminating them completely could have unintended consequences. In particular, AG opinions may serve as a public signal of the merits of the case, which might deter the Court from making overtly political rulings (Carrubba & Gabel, 2015). The recent trend of using AG opinions in cases that present novel legal issues seems like a reasonable approach.

Our simulation predicts that using 3-judge panels would have had a significantly smaller impact, as the Court already hears most cases in 3-judge or 5-judge chambers. Moreover, there are important side effects to consider. When there is substantial variation in the preferences of judges, using smaller chambers will increase the variance of judicial outcomes, decreasing the consistency of EU law (Fjelstul, 2022).

Future research can look for opportunities to leverage causal inference techniques to precisely identify the causal effects of historical reforms on the Court's productivity. There are two reforms, in particular, that stand out: the creation of the General Court in 1989 and the expansion of the General Court, starting in 2016. If possible, precisely identifying the causal effects of these major institutional reforms would also help policy-makers evaluate future reforms to the Court of Justice.

In addition, future research can apply our methodology for estimating the effect of potential institutional reforms to other courts beyond the EU context. Our methodology is flexible enough that scholars can include the specific institutional features that are relevant to case duration at other domestic and international courts.

Notes

1. This quote is typically attributed to William Gladstone.
2. The CJ used to hear disputes involving EU employees, but these are currently heard by the GC.
3. There is not enough empirical variation to estimate the effect of the number of référendaires independent of the number of judges.
4. We estimate the model in `Stan` using `brms` in R (Bürkner, 2017; Carpenter et al., 2017). We do Markov chain Monte Carlo (MCMC) sampling of the posterior distribution using the NUTS algorithm (Hoffman & Gelman, 2014). We estimate 4 MCMC chains with 4,000 iterations per chain, including burn-in period of 2,000 iterations. We use weakly informative priors.
5. The Court's IT expenditures are contained in budget items 2100 (Purchase, servicing and maintenance of equipment and software) and 2102 (External services for the operation, creation and maintenance of software and systems). Prior to 1998, IT expenditures were contained in Chapter 22.
6. See document 32020B0227 in EUR-Lex.
7. See documents 32018B0251 and 32020B0227 in EUR-Lex.

Acknowledgments

We thank Sivaram Cheruvu, Daniel Naurin, Michal Ovádek, Øyvind Stiansen, the editor, the anonymous reviewers, and the participants at the IUROPA Workshop in Gothenburg, Sweden in November 2021 for their helpful comments and suggestions.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Joshua C. Fjelstul is a Post-Doctoral Research Fellow in the Department of Political Science and International Relations at the University of Geneva, Switzerland, and a Researcher at the ARENA Centre for European Studies at the University of Oslo, Norway.

Matthew Gabel is a Professor of Political Science in the Department of Political Science at Washington University in St. Louis, USA.

Clifford J. Carrubba is a Professor of Political Science and the Chair of the Department of Quantitative Theory and Methods at Emory University, USA.

ORCID

Joshua C. Fjelstul  <http://orcid.org/0000-0002-9494-260X>

References

- 255 Comité (2019). Sixth activity report of the panel provided for by Article 255 of the Treaty on the Functioning of the European Union.
- Brekke, S. A., Fjelstul, J. C., Hermansen, S. S. L., & Naurin, D. (2022). The CJEU database platform: Decisions and decision-makers. *Journal of Law and Courts*. <https://doi.org/10.7910/DVN/HURMWM>.
- Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1), 1–28. <https://doi.org/10.18637/jss.v080.i01>
- Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, 71(1), 1–32. <https://doi.org/10.18637/jss.v076.i01>
- Carrubba, C. J., & Gabel, M. J. (2015). *International Courts and the Performance of International Agreements*. Cambridge University Press.
- Cheruvu, S. (2019). How do institutional constraints affect judicial decision-making? The European Court of Justice's French language mandate. *European Union Politics*, 20(4), 562–583. <https://doi.org/10.1177/1465116519859428>
- Cheruvu, S., & Fjelstul, J. C. (2021). Improving the efficiency of pretrial bargaining in disputes over noncompliance with international law: Encouraging evidence from the European Union. *Journal of European Public Policy*, 29(8), 1249–1267. <https://doi.org/10.1080/13501763.2021.1945130>
- Court of Justice of the European Union (2021). The year in review: Annual review 2020.
- Dashwood, A., & Angus, J. (2001). *The Future of the Judicial System of the European Union*. Hart Publishing.
- European Court of Auditors (2017). Performance review of case management at the Court of Justice of the European Union.
- European Court of Justice (1995). Report of proceedings, 1992–1994.
- Fjelstul, J. C. (2019). The evolution of European Union law: A new data set on the Acquis Communautaire. *European Union Politics*, 20(4), 670–691. <https://doi.org/10.1177/1465116519842947>
- Fjelstul, J. C. (2022). How the chamber system at the CJEU undermines the consistency of the court's application of EU law. *Journal of Law and Courts*. <https://doi.org/10.1086/717422>.
- Frankenreiter, J. (2017). The politics of citations at the ECJ—policy preferences of E.U. member state governments and the citation behavior of judges at the European Court of Justice. *Journal of Empirical Legal Studies*, 14(4), 813–857. <https://doi.org/10.1111/jels.2017.14.issue-4>
- Frankenreiter, J. (2018). Are advocates general political? Policy preferences of EU member state governments and the voting behavior of members of the European Court of Justice. *Review of Law and Economics*, 14(1). <https://doi.org/10.1515/rle-2016-0037>
- Gelman, A., Goodrich, B., Gabry, J., & Vehtari, A. (2019). R-squared for Bayesian regression models. *The American Statistician*, 73(3), 307–309. <https://doi.org/10.1080/00031305.2018.1549100>

- Hermansen, S. S. L. (2020). Building legitimacy: Strategic case allocations in the Court of Justice of the European Union. *Journal of European Public Policy*, 27(8), 1215–1235. <https://doi.org/10.1080/13501763.2020.1714697>
- Hoffman, M. D., & Gelman, A. (2014). The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15(1), 1593–1623.
- Kelemen, R. D. (2012). The political foundations of judicial independence in the European Union. *Journal of European Public Policy*, 19(1), 43–58. <https://doi.org/10.1080/13501763.2012.632127>
- Kenney, S. J. (2000). Beyond principals and agents: Seeing courts as organizations by comparing Référendaires at the European Court of Justice and law clerks at the U.S supreme court. *Comparative Political Studies*, 33(5), 593–625. <https://doi.org/10.1177/0010414000033005002>
- Lenaerts, K., Maselis, I., & Gutman, K. (2014). *EU Procedural Law*. Oxford University Press.
- Malecki, M. (2012). Do ECJ judges all speak with the same voice? Evidence of divergent preferences from the judgments of chambers. *Journal of European Public Policy*, 19(1), 59–75. <https://doi.org/10.1080/13501763.2012.632143>
- Strasser, S. (1995). Evolution and effort: Docket control and preliminary references in the European Court of Justice. *Columbia Journal of European Law*, 2, 49–105.
- Yeung, T. W.-C., Ovádek, M., & Lampach, N. (2022). Time efficiency as a measure of court performance: Evidence from the Court of Justice of the European Union. *European Journal of Law and Economics*, 53, 209–234. <https://doi.org/10.1007/s10657-021-09722-5>