

The Czech Constitutional Court Dataset*

ANONYMIZED, ANONYMIZED

The article at hand presents a Czech Constitutional Court (“CCC”) dataset, a comprehensive high-quality multi-user dataset on the CCC. The dataset encompasses all decisions of the CCC starting from its founding in 1993 until the end of 2023, including plenty of metadata - such as information on the judge rapporteur, subject matter, or concerned legal acts, a complete text corpus, as well additional background information on judges or clerks. The lack of high-quality datasets has until recently arguably hindered empirical legal research in the European context. While that has in the last couple of years significantly changed for the better, the Central and Eastern European region still remains rather untouched by this development. Therefore, the dataset at hand is one of the first, if not the first dataset in the CEE region. To this end, it can kickstart and assist similar efforts elsewhere in the region. The CCC dataset unlocks potential for methodologically rigorous research in a region that has been gaining on attention and importance in the European context.

1 Introduction

Although it has been traditionally espoused that there has been a divide between the empirically oriented US legal scholarship, stemming from a different perception of the role of courts and judges, and the rest of the world (Hamann 2019, 416). Therein the judges empirically researched whether and to what extent they behave as for example political or strategic actors. (Carrubba et al. 2012; Clark and Lauderdale 2010; Epstein and Knight 1997, 2000; Lauderdale and Clark 2014; Sunstein et al. 2006; Cameron and Kornhauser 2017; Clark, Engst, and Staton 2018; Epstein, Landes, and Posner 2011; Kornhauser 1992b, 1992a; Posner 1993, 2010; Roussey and Soubeyran 2018).

In contrast to, especially in European legal systems, such as the one at hand - Czechia, judges have been perceived as “proclaimers of law” and the law handed down by them. (Hamann 2019, 417) even claims that such a view had hindered empirical legal research in Europe. The lack of empirical legal research can be partially blamed on lack of high quality data, a prerequisite for any quantitative empirical research. At least so the story goes until recently. The interest in empirical legal studies has picked up in the last years across the whole continent, including studies on plethora of topics within Germany (Arnold, Engst, and Gschwend 2023; Coupette and Fleckner 2018; Benjamin G. Engst et al. 2017; Wittig 2016), Spain and Portugal (Hanretty 2012), the UK (Hanretty 2020) or the EU institutions (Bielen et al. 2018; Brekke, Naurin, et al. 2023; Fjølseth 2023, 2019; Fjølseth, Gabel, and Carrubba 2022).

Publications of new high quality publicly accessible data have gone hand in hand with these developments. To the many released comprehensive data sets in the recent years belong namely the Europa project’s CJEU database (Brekke, Fjølseth, et al. 2023), the German Federal courts (Hamann 2019) as well as the German Federal Constitutional Court (Benjamin G. Engst, Hönnige, and Gschwend Forthcoming). The mushrooming research proves that there is demand for quality data in Europe as well. The article at hand presents a Czech Constitutional Court (“CCC”) dataset, a comprehensive high-quality multi-user dataset on the CCC.

The CCC dataset is foundational in that it encompasses plethora of data that other researches can base their research efforts on, it has the capacity address research questions and it adheres

*Replication files are available on the author’s Github account (ANONYMIZED). **Current version:** January 30, 2024

to the tidy data principles. The dataset encompasses all decisions of the CCC starting from its founding in 1993 until the end of 2023, including plenty of metadata - such as information on the judge rapporteur, subject matter, or concerned legal acts, a complete text corpus, as well additional background information on judges or clerks.

To the my best knowledge, the CCC dataset is one of the first, if not the first, comprehensive dataset coming out of the Central and Eastern European (“CEE”) region. The CEE region has gotten to the spotlight of the European legal research, for example, thanks to the various rule of law crises, in which the regional constitutional courts and their interplay with the CJEU have played an important role (R. Daniel Kelemen 2020; R. Daniel Kelemen and Pech 2019; Sadurski 2019 and many other articles). Despite that the CEE scholarship has so far produced very little in terms of methodologically rigorous empirical legal research output concerning the role of the judiciary, constitutional courts or judicial politics. The lack of high-quality datasets is undoubtedly a piece of this puzzle.

When we zero in on Czechia, there have been solitary attempts go gather data in some shape or form in the CCC context (Harašta et al. 2018; Novotná and Harašta 2019), mainly thanks to the Institute of Law and Technology based in Brno, as well as isolated attempts to conduct network analysis or research employing natural language processing and alike methods (Chmel 2017; Eliášek, Kól, and Švaňa 2020; Harašta et al. 2021; Vartazaryan 2022). Unfortunately, the former group does not always adhere to the principles of high-quality infrastructure, namely the principle of foundationality, espoused by Weinshall and Epstein (2020), p. 424, the latter group did not publish data/code at all.

Therefore, the effort to put together and to publish a high-quality dataset on the CCC is more than warranted, especially to enable empirical legal scholarship to flourish. As the article explains later, the dataset at hand can serve as an foundation for wide variety of research inquiries, for example, research into the role of gender or the role of law clerks. Moreover, the dataset offers a blueprint for future efforts to build akin datasets. In building the dataset, I attempted to name the variables and to structure the data in a transparent, replicable and comparable way so that any efforts from different courts can mimic my approach without steep costs and efforts.

The article proceeds as follow. In the section 2, I introduce the CCC, namely its compositions, its internal organization and its powers to give the reader a little bit of context. In the section 3, I introduce the CCC dataset. Therein, I briefly discuss its structure, its creation, and describe its variables. The section 4 then discusses the adherence of the CCC dataset to four principles of a high-quality dataset, including its relevance for research, as well as to the adherence of the tidy data principles. The last section concludes.

2 A Brief Primer on the CCC

The CCC consists of fifteen justices, out of which one is the president of the CCC, two are vice presidents and twelve associate justices (following the terminology of Kosař and Vyhnánek 2020). These fifteen justices are appointed by the president of the Czech republic upon approval of the Senate. The justices enjoy 10 years terms with the possibility of re-election; there is no limit on the times a justice can be re-elected. The three CCC functionaries are unilaterally appointed by the Czech president.

Regarding the competences, the CCC is a typical Kelsenian court inspired mainly by the German Federal Constitutional Court. The CCC enjoys the power of abstract constitutional review,

including constitutional amendments. The abstract review procedure is initiated by political actors (for example MPs) and usually concerns political issues. Moreover, an ordinary court can initiate a concrete review procedure, if that court reaches the conclusion that a legal norm upon which its decision depends is not compatible with the constitution. Individuals can also lodge constitutional complaints before the CCC. Lastly, the CCC can also resolve separation-of-powers disputes, it can *ex ante* review international treaties, decide on impeachment of the president of the republic, and it has additional ancillary powers (for a complete overview, see [Kosař and Vyhnaněk 2020](#)).

The CCC is an example of a collegial court. From the perspective of the inner, the CCC can decide in four bodies: (1) individual justices, (2) 3-member chambers (*senáty*), (3) the plenum (*plénium*), and (4) special disciplinary chambers. However, the 3-member chambers and the plenum play a crucial role. The plenum is composed of all justices, whereas the four 3-member chambers are composed of the associate justices. Neither the president of the CCC or her vice-presidents are permanent members of the 3-member chambers. Until 2016, the composition of the chambers was static. However, in 2016, a system of regular 2-yearly rotations was introduced, wherein the president of the chamber rotates to a different one every 2 years. I am of the view that such a institutional change opens up potential for quasi-experimental research similar to the Gschwend, Sternberg, and Zittlau ([2016](#)) study utilizing judge absences within the 3-member panels of the German federal constitutional court. In general, the plenum is responsible for the abstract review, whereas the 3-member chambers are responsible for the individual constitutional complaints.

In the chamber proceedings, decisions on admissibility must be unanimous, decisions on merits need not be, therefore, two votes are necessary.¹ In the plenum, the general voting quorum is a simple majority and the plenum is quorate when there are ten justices present. The abstract review is one of the exceptions that sets the quorum higher, more specifically to 9 votes.

A judge rapporteur plays a crucial role ([Chmel 2017](#); [Hořeňovský and Chmel 2015](#) study the large influence of the judge rapporteurs at the CCC). Each case of the CCC gets assigned to a judge rapporteur. The assignment is regulated by a case allocation plan.² She is tasked with drafting the opinion, about which the body then votes. The president of the CCC (in plenary cases) or the president of the chamber (in chamber cases) may re-assign a case to a different judge rapporteur if the draft opinion by the original judge rapporteur did not receive a majority of votes. Unfortunately, the CCC does not keep track of these reassignments.³

The CCC allows for separate opinions. They can take two forms: dissenting or concurring opinions. Each justice has the right to author a separate opinion, which then gets published with the CCC decision. It follows that not every anti-majority vote implies a separate opinion, it is up to the justices to decide whether they want to attach a separate opinion with their vote. Vice-versa, not every separate opinion implies an anti-majority vote, as the justices can attach a concurring opinion. In contrast to dissenting opinion, when a judge attaches a concurring opinion, they voted with the majority but disagree with its argumentation.⁴

¹Which enables the attachment of separate opinions

²the original term is *rozvrh práce*, which is usually translated as a *work schedule*, however, I borrow the term *case allocation plan* from Hamann ([2019](#)), p. 673

³More specifically, some decisions mention the reassignment. However, not all do, therefore an attempt to retrieve the information from the texts was highly unreliable. According to my internal insight, the CCC should always pass a short procedural decision when a reassignment takes place. I unsuccessfully attempted to retrieve the information with the right to information as the procedural decision is not available in electronic form and retrieving the full information would thus entail manually going through all paper files.

⁴Which makes it difficult to, for example, conduct the same point-estimation with data on dissenting behavior of justices as Hanretty ([2012](#)) has done on the Portuguese and Spanish Constitutional Courts.

It may be concluded that the CCC takes after the American model of selection of justices, with the president of the republic and the upper chamber being in the spotlight, but is also a typical example of a Kelsenian specialised court with concentrated constitutional review. The CCC stands out in how powerful its constitutional review is, having attracted the power to review even constitutional amendments, thus, the CCC is a powerful player in the Czech political system. The internal organization of the CCC gives a lot of room for strategic considerations of its justices. Not only due to the similarities with the constitutional adjudication powerhouses but also due to its own idiosyncrasies, I believe the CCC to be a worthy object of empirical legal research.

3 Description of the CCC Dataset

3.1 Structure and Case Inclusion

The CCC dataset includes all publicly available CCC decisions from its foundation until the end of 2023, that is 93826 decisions, as well as background information on its 50 justices and their 221 clerks. All the data was first scraped from the official CCC website [Nalus](#). The web scraping is followed by intense data cleaning and data wrangling processes. In the last step, some information was mined from the texts or other metadata of the decisions (such as composition of the court in any given decision). The CCC dataset is accompanied by a comprehensive codebook, which contains detailed explanation of its structure, parts and all variables contained therein.⁵

The structure of the CCC dataset can, on a very basic level, be divided into decision-level variables and justice-level/clerk-level variables. As I explain below such a structure does not imply that on the decision-level, for example, every row is a decision. The decision (as well as the justice-level) variables have been split up into multiple datasets for those variables that would otherwise imply more information within one cell, had the justice level structure been kept. For example, the composition of the court is a decision-level variable - each chamber decision has certain composition of three justices - however, in order to keep the one observation (decision) per row structure, one would have to store multiple information in one cell (names of three or fifteen justices). Therefore, while I so flatten and nest the data for the “master” metadata dataset, for each variable there is also a separate table, in which the level is the decision-variable (so for example a decision-sitting_judge).

3.2 Decision-level Variables

The identifying decision variables contain information on each and every case that the CCC has decided. There are variables uniquely identifying the decisions (ECLI renamed to doc_id or case_id identifying the whole case, which can contain more than one decisions) or the dates at which they were lodged and decided.

The procedural variables concern whether the decision was a *usnesení* or *nález* (roughly speaking *nález* is a decision on merits, whereas *usnesení* is a decision on admissibility or a procedural decision), what type of procedure the decision was made in (abstract review, constitutional complaint etc.), or on what type of grounds the decision was based.

The background variables concern among others parties before the CCC, which are identified (a natural person, a legal person, a court etc.), the body whose decision was under review (typically which court), the type of decision being review and alike. Moreover, the data on the subject proceedings (relates to the area of constitutional law) and subject register (the pertaining area

⁵Available at: XXX

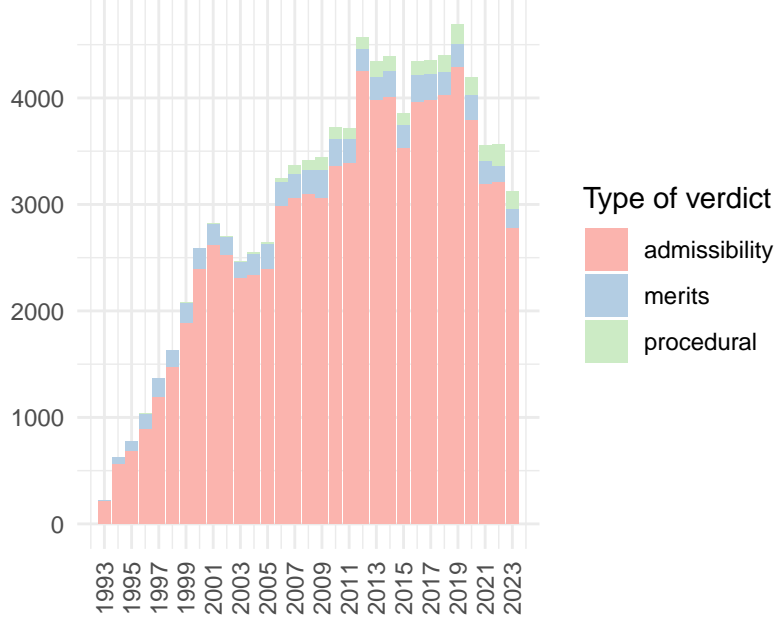


Figure 1: We can see the developing caseload overtime with the colors referring to the type of the decision. We can observe that while the caseload of the CCC has grown, the proportions between the admissibility and merits decisions have not changed, despite the efforts of the CCC to administer more decisions more quickly with the decisions on admissibility (especially the ill-foundness).

of general law such as criminal-proof, civil damages, or administrative proceedings) variables are included. Such variables are especially useful for controlling for specific features of cases that may have confounding potential.

Lastly, the CCC dataset always contains the full text of the decision, which unlocks plethora of potential research utilizing quantitative text analysis or various machine learning endeavor. A number of variables have already been mined from these texts. To name two, the compositions of the chambers have been mined using various regex variations of the justices’ names and dissenting opinions as well as their relationships to each other (whether more judges signed one dissenting opinion or whether they dissented separately) have been mined from the texts.

3.3 Justice-Level Variables

Justice-level and clerk-level variables contain information on the individual justices and clerks respectively. Most of the information was collected manually from the official profile of the justices at the CCC website.

A lot of information about the justices can be learned from the decision: which justice was the judge rapporteur in any given case, which justice sat on which case, or in which case and with whom did any given justice dissent. The purely justice-level variables concern two main groups of information: biographic and about their clerk team.

Regarding the former, the CCC dataset includes information on the terms of the justices, their age and gender, their alma mater, highest reached degree, as titles play an especially important “ceremonial” role in the Czech legal environment, their professional background before they became a judge, or information on whether the justice ran for a reelection.



Figure 2: The histogram plot of the age of a justice at the start of their term is bimodal. One peak occurs roughly in the justices’ early-fifties, which is quite reasonable, whereas the other occurs in their early-sixties. That is quite late given that the retirement age of ordinary judges is sixty-five and given that the term of CCC justices is 10 years long. Therefore, many justices are serving well into their seventies.

We can, for example observe, that although the CCC has a seat in Brno, most of its justices have graduated at the Prague Charles University. One could, for example, study whether the norm-identification with the norms surrounding their role of a CCC justice as presented by Wittig (2016) varies among the justices based on their *alma mater*, as the judges from Prague must have either relocated to a new city or commuted to their work⁶, whereas the graduates from Brno’s Masaryk University have been socialized into the CCC during their studies, as many of the CCC justices have historically kept their teaching or research duties at MUNI, and do not have to commute.

Secondly, the dataset also includes information on all 221 clerks that have served in the CCC’s history. To this end, I believe inclusion of such an information makes the dataset quite unique and opens up a lot of avenues for research. The information on the clerks contains under which judge they deserved, what was their term, what is their gender, education, or whether they studied abroad.

⁶Based on a personal insight, there is a number of examples of the later approach, namely of justices that never identified themselves with Brno and that would commute to the CCC only for a day or two per week.

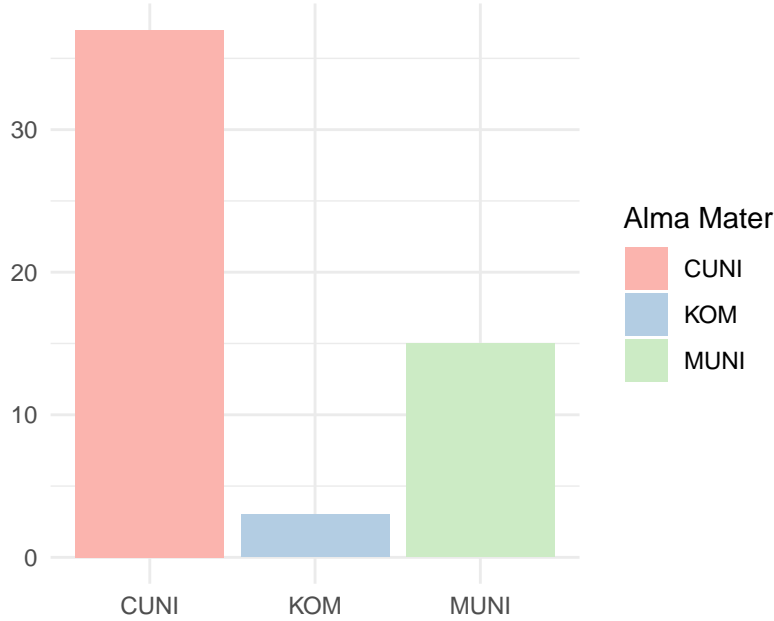


Figure 3: The alma mater among the CCC justices.

4 Principles Guiding the CCC Dataset

The Czech Constitutional Court Dataset is a “multi-user dataset” created in a principled manner. Epstein et al. (2014), p. 14 define a multi-user dataset as a dataset created with the purpose of that “[r]ather than collect data to answer particular research questions [...] the idea is to amass a dataset so rich in content that multiple users, even those with distinct projects, can draw on it.”

Accordingly, the Czech Constitutional Court dataset upholds the principles of a high-quality dataset espoused by Weinshall and Epstein (2020), p. 424, namely that the database is (1) capable of addressing real-world problems, (2) accessible, (3) reproducible and reliable, and (4) foundational.⁷ The data structure also follows the principles of tidy data. According to Wickham (2014), tidy data are data with such a tabular⁸ structure that stick to the following principles

- (1) every column is a variable,
- (2) every row is an observation,
- (3) every cell is a single value.

Although the dataset contains one all-encompassing metadata dataset, some of its variables (for example concerned acts or applicant) contain multiple values nested in one cell. The reason is simple: the metadata dataset contains observations on the decision, whereas whenever a variable contains more values per case, the data structure would then require a variable-case level. For this purpose, the dataset is also split up into multiple smaller datasets on a variable-case level (for example dissenting judge-case level or concerned acts-case level), which can then be joined together relational database SQL-style in the form of unique keys.

⁷I decided to skip the principle of sustainability as I do not see the benefit of discussing it separately, most of its issues are answered in the other sections

⁸i.e. data with a column and row structure

4.1 Capacity to Address Real-World Problems

In the words of Weinshall and Epstein (2020), “By definition, data infrastructure should promote innovation, inventions, and insights. Although no product can guarantee these ends, infrastructure aimed at solving (or developing implications for) real-world problems increases the odds of success.” With the dataset at hand I hope to enable data- and evidence- based research on the CCC, which has so far not been an object of thorough empirical legal research. I now present two examples that corroborate the capacity of the CCC dataset to address real-world problems and research concerns.

4.1.1 Clerks

The first brief example concerns clerks of justices. Kosař and Vyhnánek (2020) argue that the clerks at the CCC play an especially vital and unappreciated role: “The initial idea of the legislature was to grant each Justice one law clerk who would take administrative burdens unrelated to substantive decision-making off the Justices’ shoulders. Yet the reality is different. First, due to the growing caseload, the number of law clerks per Justice increased gradually; today, each Justice has three law clerks. Moreover, law clerks de facto prepare drafts of most CCC judgments and decisions, and the real administrative burden has been ‘outsourced’ to secretaries of the cabinets.” The difficulty of studying the role of clerks was highlighted in the Clark, Engst, and Staton (2018) study on the effects of leisure on judicial performance.

The CCC dataset contains the information on all clerks that have served on the CCC, including their gender, education, beginning and end of mandates. Such an information can serve as a basis for any research on the role of clerks. For the purpose of showing the capability of solving real-world problems, I present descriptive statistics. We can for example compare the gender distribution of clerks against that of judges. While we observe almost complete gender equality among the clerks

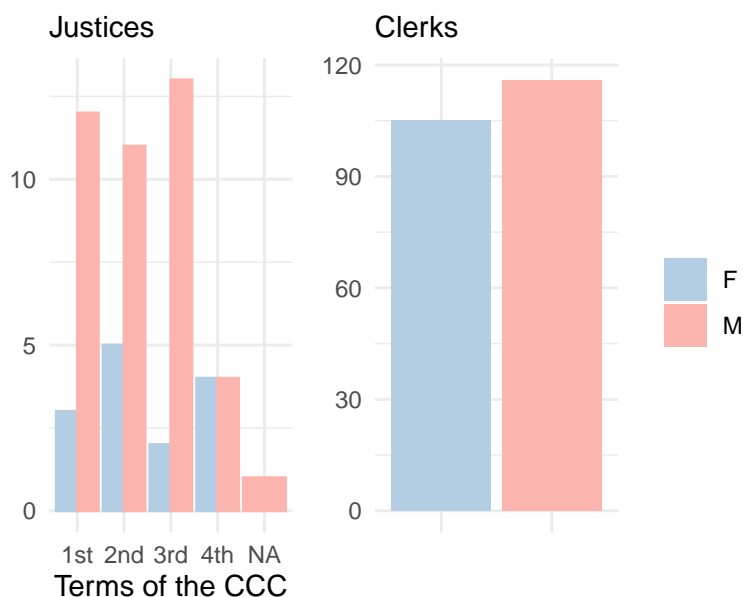


Figure 4: Comparison of proportions of genders among justices (on the left) and among their clerks (on the right). The X axis of the left bar chart signifies the terms of the CCC, which roughly correspond to decades.

We can further inquire whether male or female justice have different preferences regarding their

clerks. Surprisingly, given the theoretical expectations, the judges at the CCC seem to prefer clerks equally whereas female judges seem to hire more male clerks.



Figure 5: The different gender composition of the clerk teams faceted by the gender of the hiring justice. The gender composition of the clerk teams of female justices is on the left and their male counterparts are on the right.

The data at hand have a potential to shed light on the understudied role of law clerks, whether its their influence on the final decision as “an information source” (Kromphardt 2015) or the influence of their background (such as gender) on their behavior (Badas and Stauffer 2023). Such a research is far and few between even in the US context and it’s practically non-existent in the European context.

4.1.2 Dissenting behavior of justices

The research on judicial coalitions at the CCC has revealed that the third period of CCC between 2013-2023 is rather polarized and that there are two big coalitions of judges that clash against each other (Chmel 2021; Smekal et al. 2021; Vartazaryan 2022). The articles rely primarily on network analysis of the dissenting opinions in the plenary proceedings and make inferential conclusions based on a rather superficial descriptive analysis. I predict that should the relationship from the plenum indeed exist, they should also carry over to the 3-member panel hearings. The hypothesis is that panels composed of judges from both coalitions will be more likely to show disagreement in the form of dissenting opinion; i.e. whether having a 3-member panel composed of members of both judicial coalitions increases judges’ likelihood of a dissent. If this shows to be true, it would provide further evidence to the two coalition theory of the CCC (Chmel 2021; Vartazaryan 2022; Smekal et al. 2021).

I very concisely tested whether the presumable existence of the coalitions carries over to and has any effect on the dissenting behavior of judges in the panels building a very simple logistic regression. I manually annotated which justices of the 3rd term were from which coalition, thus further proving that the CCC dataset adheres to the below discussed principle of foundationality

Table 1: Results from the Logit Model

	(1)
(Intercept)	−3.783*** (0.125)
coalitionfull_coal	−1.214*** (0.316)
Num.Obs.	4710
AIC	771.3
BIC	784.2
Log.Lik.	−383.645
RMSE	0.13

as it can be additionally built upon.

The model predicts the likelihood of a judge attaching a separate opinion occurring in the 3-member panel proceedings depending on whether the panel is fully composed of members from the 1st or 2nd coalition and on the composition being a mix of one minority justice from one coalition and the remaining two from the other. The dependent variable is thus a dummy variable containing the information whether a judge attached a separate opinion or not and the independent variable is a categorical variable of whether the panel was composed of members of one coalition or whether the panel was mixed.

The very simple model reveals that there does seem to be a negative association between the likelihood of a judge to dissent and the panel being composed fully of members from one or the other coalition. Such a conclusion is in line with the Czech legal scholarship and warrants further more in-depth inquiry.

4.1.3 Natural language processing

The last potential usage of the CCC dataset that immediately springs into mind is the application of various natural language processing methods. For example, one could replicate the research on vagueness of language of the CCC (Sternberg 2019) or measure the readability of the CCC decisions (Crossley, Skalicky, and Dascalu 2019; Fix and Fairbanks 2020) and link those measures to interesting research question. For example, do the better readable CCC decisions get cited by the CCC more than the less readable (Fix and Fairbanks 2020; Crossley, Skalicky, and Dascalu 2019)? Does the CCC use vague language in certain areas/for certain reasons more than for others (Sternberg 2019)?

Practically every dataset introduction article contains a disclaimer that the goal of presenting these simplified models is not to draw any inference but rather to show the potential of using the dataset “to develop real-world implications and contribute to public and academic discourse on pressing legal-political issues”. (Weinshall and Epstein 2020, 427) This section has attempted to do exactly that.

4.2 Accessibility

The principle of accessibility demands that “in the creation of high-quality infrastructure is that members of the community should be able to access it with no barriers to entry or use.” (Weinshall and Epstein 2020, 427)

As I have shown in the introduction with specific examples, not all research is reproducible and not all data are made available. That goes against the principle of accessibility. Weinshall and Epstein refer to studies, according to which majority of psychological research data stays under embargo or never gets released at all (Houtkoop et al. 2018) or that only minority of papers published in journals requiring a data availability statement actually do publish their data (Federer et al. 2018). There is no need to prove why such a practice is wrong, why it hinders reproducibility of science.

Following the principle of accessibility, the Czech Constitutional Court Dataset is freely and publicly available in full, with the handbook as well as this article attached to it. **That data are downloadable at** ⁹ The data are published out of my own accord, the publication is not funded by any grant or national science foundation.

4.3 Reliability and Reproducibility

Weinshall and Epstein (2020) define the principles as follows: “Reproducibility means that users and developers alike must understand how to duplicate the data housed in the infrastructure. Reliability is related: it is the extent to which encoded data can be replicated, producing the same value using the same standard for the same subject at the same time, regardless of who or what is doing the replicating” The heart of the matter of reliability and reproducibility is internal consistency of the dataset, not necessarily its external validity.

The data must have been reliably generated. In my case, I did not narrow down the selection of cases - all cases of the CCC that have been made publicly available from its history have been web scraped from its website, including all the available information as well as the texts of the decisions. Reproducibility also demands that anyone with sufficient skill should be able to reproduce the dataset on their own based on the provided information. All the code has been made available on Github, the code is written in a clean manner and is well commented.

Bound to both principles is the issue of coding the variables. To this end, the amount of human input has been minimized. Vast majority of the information provided has either been directly (or with minimal input) collected from the CCC website or it has been automated to the maximal possible extent (including the full information on the clerks). Only the biographic information on judges has been imputed via human input using the official profiles of justices at the CCC website as well as Wikipedia as sources. The rest is product of the published code.

Therefore, the CCC dataset is highly reliable and easily reproducible, that is at least with enough time to run the whole web scraping process.

4.4 Foundational

The principle that a dataset be foundational requires that it should serve “as a foundation upon which researchers can build by adding content, backdating, updating, or otherwise adapting it to their own needs; it should not be the be-all, end-all.” In other words, the principle promotes a

⁹ADD LINK WHEN PUBLISHED

generally usable data over one-off solutions to particular research questions.

The CCC dataset is foundational. The dataset includes comprehensive background data on each and every case, bibliographic data on the justices, quite unique data on the clerks, as well as a full text corpus of all the decisions.

The CCC dataset can thus serve researches interested in judicial behavior, in institutional studies, or anyone wanting to leverage natural language processing on large amount of text data. I am currently working on a paper exploring the citations of CCC caselaw to each other to estimate the location of cases in the doctrine space, as well as on paper leveraging machine learning classification algorithm on the text corpus to study dissenting behavior of the CCC justices. Therefore, it has already proven foundational, notwithstanding the research question I have come up with.

5 Conclusion

The present article introduced a dataset on the Czech Constitutional Court while aiming at bridging the gap. To this end, the dataset contains rich information on the background of the cases, background of the justices, their clerk teams as well as the complete text corpus of all decisions until the end of 2023. The dataset adheres to the principles of tidy data as well as to the principles of accessibility, reliability, foundationality, and capacity to address real-world problems.

The dataset, in my view, enables empirical research in the CEE region that has been lacking in the past on methodologically rigorous empirical research. The datasets unlocks research on the decision-making of judges and judicial politics, such as their dissenting behavior, strategic acting, or the influence of their clerk teams, on the institutional set up of the CCC, such as the introduction of rotations or the various ways to expedite the CCC caseload, and lastly on the texts of the decisions themselves, for example studying various linguistic features of the decisions such as readability or vagueness. Therefore, the present article makes a valuable contribution to the (European) empirical legal research scholarship.

Literature

- Arnold, Christian, Benjamin G. Engst, and Thomas Gschwend. 2023. “Scaling Court Decisions with Citation Networks.” *Journal of Law and Courts* 11 (1): 25–44. <https://doi.org/10.1086/717420>.
- Badas, Alex, and Katelyn E. Stauffer. 2023. “Gender and Ambition Among Potential Law Clerks.” *Journal of Law and Courts* 11 (1): 116–40. <https://doi.org/10.1017/jlc.2022.8>.
- Bielen, Samantha, Ludo Peeters, Wim Marneffe, and Lode Vereeck. 2018. “Backlogs and Litigation Rates: Testing Congestion Equilibrium Across European Judiciaries.” *International Review of Law and Economics* 53 (March): 9–22. <https://doi.org/10.1016/j.irle.2017.09.002>.
- Brekke, Stein Arne, Joshua C. Fjelstul, Silje Synnøve Lyder Hermansen, and Daniel Naurin. 2023. “The CJEU Database Platform: Decisions and Decision-Makers.” *Journal of Law and Courts*, January, 1–22. <https://doi.org/10.1017/jlc.2022.3>.
- Brekke, Stein Arne, Daniel Naurin, Urška Šadl, and Lucía López-Zurita. 2023. “That’s an Order! How the Quest for Efficiency Is Transforming Judicial Cooperation in Europe.” *JCMS: Journal of Common Market Studies* 61 (1): 58–75. <https://doi.org/10.1111/jcms.13346>.

- Cameron, Charles M., and Lewis A. Kornhauser. 2017. "Chapter 3: What Do Judges Want? How to Model Judicial Preferences." SSRN Scholarly Paper. Rochester, NY. June 2, 2017. <https://doi.org/10.2139/ssrn.2979419>.
- Carrubba, Cliff, Barry Friedman, Andrew D. Martin, and Georg Vanberg. 2012. "Who Controls the Content of Supreme Court Opinions?" *American Journal of Political Science* 56 (2): 400–412. <https://doi.org/10.1111/j.1540-5907.2011.00557.x>.
- Chmel, Jan. 2017. "Zpravodajové a Senáty: Vliv Složení Senátu Na Rozhodování Ústavního Soudu České Republiky o Ústavních Stížnostech." *Časopis Pro Právní Vědu a Praxi* 25 (4): 739. <https://doi.org/10.5817/CPVP2017-4-9>.
- . 2021. *Co Ovlivňuje Ústavní Soud a Jeho Soudce?* /. Vydání první. Teoretik (Leges). Leges,.
- Clark, Tom S., Benjamin G. Engst, and Jeffrey K. Staton. 2018. "Estimating the Effect of Leisure on Judicial Performance." *The Journal of Legal Studies* 47 (2): 349–90. <https://doi.org/10.1086/699150>.
- Clark, Tom S., and Benjamin Lauderdale. 2010. "Locating Supreme Court Opinions in Doctrine Space." *American Journal of Political Science* 54 (4): 871–90. <https://doi.org/10.1111/j.1540-5907.2010.00470.x>.
- Coupette, Corinna, and Andreas M. Fleckner. 2018. "Quantitative Rechtswissenschaft." *Juristen-Zeitung (JZ)* 73 (8): 379–89. <https://doi.org/10.1628/jz-2018-0020>.
- Crossley, Scott A., Stephen Skalicky, and Mihai Dascalu. 2019. "Moving Beyond Classic Readability Formulas: New Methods and New Models." *Journal of Research in Reading* 42 (3-4): 541–61. <https://doi.org/10.1111/1467-9817.12283>.
- Eliášek, Martin, Jakub Kól, and Miloš Švaňa. 2020. "Automatická Klasifikace Významových Celků v Judikatuře." *Revue Pro Právo a Technologie* 11 (21): 3–20. <https://doi.org/10.5817/RPT2020-1-1>.
- Engst, Benjamin G., Thomas Gschwend, Nils Schaks, Sebastian Sternberg, and Caroline Wittig. 2017. "Zum Einfluss Der Parteinähe Auf Das Abstimmungsverhalten Der Bundesverfassungsrichter – Eine Quantitative Untersuchung." *JuristenZeitung* 72 (17): 816–26. <https://www.jstor.org/stable/44867374>.
- Engst, Benjamin G, Christoph Hönnige, and Thomas Gschwend. Forthcoming. "The Constitutional Court Database." *Working Paper*, Forthcoming, 39.
- Epstein, Lee, and Jack Knight. 1997. *The Choices Justices Make*. SAGE. https://books.google.com?id=hSnom2h2_zUC.
- . 2000. "Toward a Strategic Revolution in Judicial Politics: A Look Back, A Look Ahead." *Political Research Quarterly* 53 (3): 625–61. <https://doi.org/10.1177/106591290005300309>.
- Epstein, Lee, William M. Landes, and Richard A. Posner. 2011. "Why (And When) Judges Dissent: A Theoretical And Empirical Analysis." *Journal of Legal Analysis* 3 (1): 101–37. <https://doi.org/10.1093/jla/3.1.101>.
- Epstein, Lee, Andrew D. Martin, Lee Epstein, and Andrew D. Martin. 2014. *An Introduction to Empirical Legal Research*. Oxford, New York: Oxford University Press.
- Federer, Lisa M., Christopher W. Belter, Douglas J. Joubert, Alicia Livinski, Ya-Ling Lu, Lissa N. Snyders, and Holly Thompson. 2018. "Data Sharing in PLOS ONE: An Analysis of Data Availability Statements." *PLOS ONE* 13 (5): e0194768. <https://doi.org/10.1371/journal.pone.0194768>.
- Fix, Michael P., and Bailey R. Fairbanks. 2020. "The Effect of Opinion Readability on the Impact

- of U.S. Supreme Court Precedents in State High Courts.” *Social Science Quarterly* 101 (2): 811–24. <https://doi.org/10.1111/ssqu.12752>.
- Fjelstul, Joshua C. 2019. “The Evolution of European Union Law: A New Data Set on the *Acquis Communautaire*.” *European Union Politics* 20 (4): 670–91. <https://doi.org/10.1177/1465116519842947>.
- . 2023. “How the Chamber System at the CJEU Undermines the Consistency of the Court’s Application of EU Law.” *Journal of Law and Courts*, 717422. <https://doi.org/10.1086/717422>.
- Fjelstul, Joshua C., Matthew Gabel, and Clifford J. Carrubba. 2022. “The Timely Administration of Justice: Using Computational Simulations to Evaluate Institutional Reforms at the CJEU.” *Journal of European Public Policy*, August, 1–22. <https://doi.org/10.1080/13501763.2022.2113115>.
- Gschwend, Thomas, Sebastian Sternberg, and Steffen Zittlau. 2016. “Are Judges Political Animals After All? Quasi-Experimental Evidence from the German Federal Constitutional Court.” SSRN Scholarly Paper. Rochester, NY. February 26, 2016. <https://doi.org/10.2139/ssrn.2738512>.
- Hamann, Hanjo. 2019. “The German Federal Courts Dataset 1950–2019: From Paper Archives to Linked Open Data.” *Journal of Empirical Legal Studies* 16 (3): 671–88. <https://doi.org/10.1111/jels.12230>.
- Hanretty, Chris. 2012. “Dissent in Iberia: The Ideal Points of Justices on the Spanish and Portuguese Constitutional Tribunals.” *European Journal of Political Research* 51 (5): 671–92. <https://doi.org/10.1111/j.1475-6765.2012.02056.x>.
- . 2020. *A Court of Specialists: Judicial Behavior on the UK Supreme Court*. Oxford University Press. <https://doi.org/10.1093/oso/9780197509234.001.0001>.
- Harašta, Jakub, Jaromír Šavelka, František Kasl, Adéla Kotková, Pavel Loutocký, Jakub Míšek, Daniela Procházková, et al. 2018. “Annotated Corpus of Czech Case Law for Reference Recognition Tasks.” In *Text, Speech, and Dialogue*, edited by Petr Sojka, Aleš Horák, Ivan Kopeček, and Karel Pala, 11107:239–50. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-00794-2_26.
- Harašta, Jakub, Terezie Smejkalová, Jaromír Šavelka, and Radim Polčák. 2021. *Citační analýza judikatury*. Vydání první. Právní monografie. Praha: Wolters Kluwer.
- Hořeňovský, Jan, and Jan Chmel. 2015. “The Process of making the Constitutional Court Judgments.” *Časopis pro právní vědu a praxi* 23 (3): 302–11. <https://www.ceeol.com/search/article-detail?id=780150>.
- Houtkoop, Bobby Lee, Chris Chambers, Malcolm Macleod, Dorothy V. M. Bishop, Thomas E. Nichols, and Eric-Jan Wagenmakers. 2018. “Data Sharing in Psychology: A Survey on Barriers and Preconditions.” *Advances in Methods and Practices in Psychological Science* 1 (1): 70–85. <https://doi.org/10.1177/2515245917751886>.
- Kelemen, R. Daniel, and Laurent Pech. 2019. “The Uses and Abuses of Constitutional Pluralism: Undermining the Rule of Law in the Name of Constitutional Identity in Hungary and Poland.” *Cambridge Yearbook of European Legal Studies* 21 (December): 59–74. <https://doi.org/10.1017/cel.2019.11>.
- Kelemen, R. Daniel. 2020. “The European Union’s Authoritarian Equilibrium.” *Journal of European Public Policy* 27 (3): 481–99. <https://doi.org/10.1080/13501763.2020.1712455>.
- Kornhauser, Lewis A. 1992a. “Modeling Collegial Courts. II. Legal Doctrine.” *Journal of Law, Economics and Organization* 8: 441. <https://heinonline.org/HOL/Page?handle=hein.journals>

- [/jleo8&id=449&div=&collection=.](#)
- . 1992b. “Modeling Collegial Courts I: Path-dependence.” *International Review of Law and Economics* 12 (2): 169–85. [https://doi.org/10.1016/0144-8188\(92\)90034-O](https://doi.org/10.1016/0144-8188(92)90034-O).
- Kosař, David, and Ladislav Vyhnánek. 2020. “The Constitutional Court of Czechia.” In *The Max Planck Handbooks in European Public Law: Volume III: Constitutional Adjudication: Institutions*, edited by Armin von Bogdandy, Peter Huber, and Christoph Grabenwarter, 0. Oxford University Press. <https://doi.org/10.1093/oso/9780198726418.003.0004>.
- Kromphardt, Christopher D. 2015. “US Supreme Court Law Clerks as Information Sources.” *Journal of Law and Courts* 3 (2): 277–304. <https://doi.org/10.1086/682136>.
- Lauderdale, Benjamin E., and Tom S. Clark. 2014. “Scaling Politically Meaningful Dimensions Using Texts and Votes: SCALING POLITICALLY MEANINGFUL DIMENSIONS.” *American Journal of Political Science* 58 (3): 754–71. <https://doi.org/10.1111/ajps.12085>.
- Novotná, Tereza, and Jakub Harašta. 2019. “The Czech Court Decisions Corpus (CzCDC): Availability as the First Step.” October 21, 2019. <http://arxiv.org/abs/1910.09513>.
- Posner, Richard A. 1993. *What Do Judges and Justices Maximize?: (The Same Thing Everyone Else Does)*. Law School, University of Chicago. <https://books.google.com?id=ciFUHQACA AJ>.
- . 2010. *How Judges Think*. Harvard University Press. <https://books.google.com?id=ZVUC8riEVPQC>.
- Roussey, Ludivine, and Raphael Soubeyran. 2018. “Overburdened Judges.” *International Review of Law and Economics* 55 (September): 21–32. <https://doi.org/10.1016/j.irl.2018.02.003>.
- Sadurski, Wojciech. 2019. *Poland’s Constitutional Breakdown*. Oxford, New York: Oxford University Press.
- Smekal, Hubert, Jaroslav Benák, Monika Hanych, Ladislav Vyhnánek, and Štěpán Janků. 2021. *Mimoprávní Vlivy Na Rozhodování Českého Ústavního Soudu*. Brno: Masaryk University Press. <https://doi.org/10.5817/CZ.MUNI.M210-9884-2021>.
- Sternberg, Sebastian. 2019. “No Public, No Power? Analyzing the Importance of Public Support for Constitutional Review with Novel Data and Machine Learning Methods.” PhD thesis, Mannheim: Universität Mannheim. <https://madoc.bib.uni-mannheim.de/52329>.
- Sunstein, Cass R., David Schkade, Lisa M. Ellman, and Andres Sawicki. 2006. *Are Judges Political? An Empirical Analysis of the Federal Judiciary*. Brookings Institution Press. <https://www.jstor.org/stable/10.7864/j.ctt12879t7>.
- Vartazaryan, Gor. 2022. “Sít’ová Analýza Disentujících Ústavních Soudců.” *Pravnik*, no. 12.
- Weinshall, Keren, and Lee Epstein. 2020. “Developing High-Quality Data Infrastructure for Legal Analytics: Introducing the Israeli Supreme Court Database.” *Journal of Empirical Legal Studies* 17 (2): 416–34. <https://doi.org/10.1111/jels.12250>.
- Wickham, Hadley. 2014. “Tidy Data.” *Journal of Statistical Software* 59 (10). <https://doi.org/10.18637/jss.v059.i10>.
- Wittig, Caroline. 2016. *The Occurrence of Separate Opinions at the Federal Constitutional Court*. Logos Verlag Berlin. <https://doi.org/10.30819/4411>.