

Hashimoto, Barry; Park, Sanghoon; Wu, Victor Y.

**Working Paper**

## Computational Reproducibility of "The Impact of Presidential Appointment of Judges: Montesquieu or the Federalists?"

I4R Discussion Paper Series, No. 135

**Provided in Cooperation with:**

The Institute for Replication (I4R)

*Suggested Citation:* Hashimoto, Barry; Park, Sanghoon; Wu, Victor Y. (2024) : Computational Reproducibility of "The Impact of Presidential Appointment of Judges: Montesquieu or the Federalists?", I4R Discussion Paper Series, No. 135, Institute for Replication (I4R), s.l.

This Version is available at:

<https://hdl.handle.net/10419/300276>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



No. 135

I4R DISCUSSION PAPER SERIES

# **Computational Reproducibility of “The Impact of Presidential Appointment of Judges: Montesquieu or the Federalists?”**

Barry Hashimoto

Sanghoon Park

Victor Y. Wu

**July 2024**

## I4R DISCUSSION PAPER SERIES

I4R DP No. 135

# Computational Reproducibility of “The Impact of Presidential Appointment of Judges: Montesquieu or the Federalists?”

**Barry Hashimoto, Sanghoon Park<sup>1</sup>, Victor Y. Wu<sup>2</sup>**

*<sup>1</sup>University of South Carolina, Columbia/USA*

*<sup>2</sup>Stanford University, Stanford/USA*

JULY 2024

Any opinions in this paper are those of the author(s) and not those of the Institute for Replication (I4R). Research published in this series may include views on policy, but I4R takes no institutional policy positions.

I4R Discussion Papers are research papers of the Institute for Replication which are widely circulated to promote replications and meta-scientific work in the social sciences. Provided in cooperation with EconStor, a service of the [ZBW – Leibniz Information Centre for Economics](#), and [RWI – Leibniz Institute for Economic Research](#), I4R Discussion Papers are among others listed in RePEc (see IDEAS, EconPapers). Complete list of all I4R DPs - downloadable for free at the I4R website.

I4R Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

### Editors

**Abel Brodeur**  
*University of Ottawa*

**Anna Dreber**  
*Stockholm School of Economics*

**Jörg Ankel-Peters**  
*RWI – Leibniz Institute for Economic Research*

# Computational Reproducibility of “The Impact of Presidential Appointment of Judges: Montesquieu or the Federalists?”<sup>\*</sup>

Barry Hashimoto<sup>1</sup>

Sanghoon Park<sup>2</sup>

Victor Y. Wu<sup>3</sup>

June 27, 2024

## Abstract

We computationally reproduce the central findings in [Mehmood \(2022\)](#), which studied the effect of a 2010 reform in Pakistan replacing the presidential appointment of high-court judges with peer appointments. Mehmood leveraged judicial records interpreted and coded by lawyers in Pakistan at the levels of cases, districts, benches, and individual judges. We successfully execute all Stata code in the author’s replication archive without any errors, then translate and execute that code in R, again finding no serious errors. Consequently, we reproduce the article’s main findings from regressions in Tables 2–4. Additionally, we successfully reconstruct the primary treatment variables of these regressions, after corresponding with the author to clarify precisely how to do so. We then replicate the main findings from regressions in Tables 2–10. Finally, we identify several minor errors which left the article’s findings intact. Overall, this report reveals no serious defects in [Mehmood \(2022\)](#). We publicly archive our replication code and a spreadsheet of our results.

**KEYWORDS:** comparative politics, judicial politics, rule of law, constitutions, replication, computational social science, Pakistan

---

<sup>\*</sup> Authors are listed in alphabetical order. None of the authors have any financial support or conflict of interest involving the original author or the original study. We thank Abel Brodeur and the Institute for Replication for assistance and feedback in the production of this report. We also thank Sultan Mehmood for his helpful replies to our inquiries. This replication was conducted during and after the Berkeley Replication Games on March 7, 2024.

<sup>1</sup>Independent. Email: [barryhashimoto@gmail.com](mailto:barryhashimoto@gmail.com)

<sup>2</sup>Political Science, University of South Carolina. Email: [SP23@email.sc.edu](mailto:SP23@email.sc.edu)

<sup>3</sup>Political Science, Stanford University. Email: [victorywu@stanford.edu](mailto:victorywu@stanford.edu)

## 1 Introduction

Mehmood (2022) investigated whether “increasing constraints on the executive—via removal of presidential discretion in judicial appointments—promotes the rule of law.” He explores this research question in the context of Pakistan’s district high courts. In 2010, Pakistan underwent a significant and unprecedented institutional overhaul that transformed its judicial selection mechanism from one where the President appointed judges for life, to a commission-based process, whereby judges are appointed by their peers.

Mehmood tested the effect of this reform on judicial decision-making, specifically pro-government rulings and decision quality. He used a reduced-reform difference-in-difference framework to compare “progovernment rulings at district benches where judges turn 62 pre-reform and are replaced by the president (control group) with district benches where judges turn 62 post reform and are replaced by judge peer appointees (treatment group).”

The main data set is a random sample of “8,500 cases—conditional on the state being one of the parties—from 1986 to 2019 for 64 high court benches (from the universe of all cases involving the government decided in this period).” Mehmood contracted two law firms (“Team 1” and “Team 2”) to code the measure of executive influence over the judiciary: a judicial dependence dummy variable of “state wins,” which took a value of one for state victories and zero for state losses. Mehmood describes his main result on pg. 426 as follows: “if 10 percent of judges retired in 2010, state wins would be about 2 percentage points lower post reform,” which is “equivalent to a 4 percent decrease over the sample mean.”

In this manuscript, we investigate the computational reproducibility of Mehmood’s results. We note that all tables and figures were computationally reproduced using the author’s replication archive by the data editor’s team at the *American Economic Journal: Applied Economics*. In other words, all numerical results produced by the Stata do-files in the replication archive match the results printed in the paper.

First, we executed the main do-file in this archive, finding no runtime errors, and verified exact replication of paper’s main results in Tables 2–4. Second, after translating the author’s Stata code to R, executing this R code successfully reproduced Mehmood’s main findings once again. Third, we initially failed to recreate the paper’s main treatment variables using the author’s replication archive, but after corresponding with Mehmood, we were able to successfully reconstruct those variables and then reproduce the main results in Tables 2–10. Finally, we report several errors in model specification and labeling, which had no material effect on the paper’s substantive results.

## **2 Dataset recreation determined to be infeasible**

We briefly considered attempting to recreate the datasets used by [Mehmood \(2022\)](#). However, we quickly determined that we would be unable to access the constituent data on Pakistani judicial records, and that it would not be feasible to perform the interpretation and coding of the data, given that Mehmood used lawyers in Pakistan to complete that task. We therefore chose to abandon this plan of recreating the datasets entirely.

## **3 Successful computational reproducibility using author’s archived Stata code**

We first executed the Stata code provided in the author’s file `master.do` of the replication archive, confirming that the tables and figures in the published paper and its appendix could be reproduced without runtime errors. Using this Stata output, we then verified that this code could reproduce the main results of the paper, which were presented in Tables 2–4. In the interest of concision, we do not document the results we obtain, which are identical to the results in [Mehmood \(2022\)](#).

In Table C20, the author replicates these main results using the second of the two data sets (i.e., the Team 2 Data Set) created by the contracted law firm and present in the replication archive. (See [Mehmood \(2022, 420\)](#) for details on the

creation of these two data sets.) We did not attempt to verify that all other results in the many tables and figures of the paper could be replicated using the Team 2 data set.

Nor, for that matter, did we verify that ancillary results appearing after those in Table 4 could be replicated using the Team 1 data set.

#### **4 Successful computational reproducibility in R**

Next, we translated the author's Stata code to R, in order to check for inadvertent errors and to better understand the control flow. Executing this R code reproduced Mehmood's main findings with near exactness.

#### **5 Initial inability to recreate the paper's main treatment variables using the author's replication archive**

In analyzing the codebase of the replication archive, we noticed that code for all of the regression models providing the paper's main results in Tables 2–4 uses three treatment variables described on page 422 paragraph 2 and page 425 paragraph 3 of [Mehmood \(2022\)](#). These plausibly exogenous variables were constructed by multiplying and dividing various other columns in the archive's data sets. Their names are:

- `Retirements_2010_Post2010`
- `appointed_bench_total2010`
- `retired_bench_total2010`

Similarly, the additional results in Tables 4, 6–10, and C2–C6, C8, and C17 use pre-made placebo variables, as mentioned on page 414, paragraph 3. These variables are named:

- `Appointment_2010_Post2010`

- `Retirements_2009_Post2010`
- `Retirements_2008_Post2010`
- `Retirements_2007_Post2010`
- `retired_sum_0009`
- `retired_sum_9099`
- `retired_sum_8689`

We hoped to replicate the paper’s regression results after recalculating these pre-generated variables ourselves. In other words, we wanted to regenerate the paper’s composite treatment variables using constituent columns in the archived data sets and then try to reproduce the paper’s results.

We were initially unable to proceed with this plan. [Mehmood \(2022\)](#) only described these variables in general terms, and he omitted the code used to make them from his replication archive. Furthermore, the relevant data sets in his archive contain a large number of variables that, according to their names, appear to record judicial retirements, appointments, and seats at various levels of analysis. We identified no transparent practice for naming these variables and could not determine what any of them recorded. Mehmood’s replication archive provides no helpful documentation in this regard, such as a codebook with descriptions and names for *all* variables in the data sets.

## **6 Successful reconstruction of the paper’s main treatment variables following correspondence with Mehmood**

With the assistance of the Institute for Replication, we sent an early draft of this report to Mehmood anonymously in late March 2024. Mehmood replied on April 10, 2024 responding to the report section above with a two-page PDF and a Stata do-file: “`data_construction.do`.” His PDF and code are included in our replication archive.



After reading these materials, we were able to straightforwardly reconstruct the treatment variables in question, as is documented in the R code of our replication archive. We then successfully reproduced nearly all of the main results in Tables 2–10 using the reconstructed treatment variables. We were only unable to reproduce one numerical estimate at this step, which we attributed to the different procedures used by R and Stata for dropping variables when the design matrix does not have full rank. Consequently, we determined that this reconstruction and replication step was a success.

## **7 Minor errors in author’s Stata code and regression model labels**

We identified several minor errors by studying the author’s Stata code and attempting to reproduce parts of it in R. None of these errors had what we would consider to be a material effect on the substantive results of the paper. We merely identify the locations of these errors.

### **7.1 Misspecification of certain fixed effects in author’s Stata code**

In Table 3 and Table C2, Mehmood reports two-stage least squares (2SLS) models with interacted district-year fixed effects (FEs). We expected covariates in both stages to be handled consistently, but in lines 533–542 and 1354–1364 of Mehmood’s Stata code, we find interacted FEs for the second stage but separate non-interacted district and year FEs for the first stage. When we re-estimated the 2SLS models using the interacted FEs at both stages, we found that the results were numerically similar (results not shown). However, this re-specification had no material effect on the substantive conclusions presented in the relevant tables.

### **7.2 Mislabeling of regression models as reported in the paper**

**7.2.1 Table 7, Panel B:** In reading and executing lines 620–627 of Mehmood’s Stata code, we discovered that the author accidentally swapped the labels identifying regression models fit to “Human rights cases” and the “Land cases” in Table 7,

Panel B reporting results for “Constitutional cases.” The label for “Human rights cases” should say “Land cases,” and vice versa.

**7.2.2 Table C8, Panel B:** In reading and executing lines 1554-1561 of Mehmood’s Stata code, we discovered that the author again swapped the labels identifying regression models fit to “Human rights cases” and the “Land cases” in Table C8, Panel B.

## 8 Conclusion

In this report, we document a successful computational reproduction of a complex article using a difference-in-difference empirical strategy published in a leading economics journal. The success of this effort was due in part to the author’s relatively transparent replication archive, as well as the rapidity and thoroughness of his reply to our initial draft replication report. In summary, while we aborted our plan to reconstruct the author’s datasets due to feasibility constraints, we were able to computationally reproduce the article’s finding by 1) executing the entire Stata codebase of Mehmood’s replication archive, 2) rewriting and running the author’s Stata code in R, and 3) reconstructing the composite treatment articles from columns in the datasets of Mehmood’s replication archive.

## References

Mehmood, S.: 2022, The impact of presidential appointment of judges: Montesquieu or the federalists?, *American Economic Journal: Applied Economics* **14**(4), 411–445.