

2021

Legal Analysis, Policy Analysis, and the Price of Deference: An Empirical Study of Mayo and Chevron

Jonathan H. Choi

University of Minnesota Law School, jonchoi@umn.edu

Follow this and additional works at: https://scholarship.law.umn.edu/faculty_articles



Part of the [Law Commons](#)

Recommended Citation

Jonathan H. Choi, *Legal Analysis, Policy Analysis, and the Price of Deference: An Empirical Study of Mayo and Chevron*, 38 YALE J. ON REG. 818 (2021), available at https://scholarship.law.umn.edu/faculty_articles/1019.

This Article is brought to you for free and open access by the University of Minnesota Law School. It has been accepted for inclusion in the Faculty Scholarship collection by an authorized administrator of the Scholarship Repository. For more information, please contact lenzx009@umn.edu.

Legal Analysis, Policy Analysis, and the Price of Deference: An Empirical Study of *Mayo* and *Chevron*

Jonathan H. Choi[†]

A huge literature contemplates the theoretical relationship between judicial deference and agency rulemaking. But relatively little empirical work has studied the actual effect of deference on how agencies draft regulations. As a result, some of the most important questions surrounding deference—whether it encourages agencies to focus on policy analysis instead of legal analysis, its relationship to procedures like notice and comment—have so far been dominated by conjecture and anecdote.

Because Chevron, U.S.A., Inc. v. Natural Resources Defense Council, Inc. applied simultaneously across agencies, it has been difficult to separate its specific causal effect from other contemporaneous events in the 1980s, like the rise of cost-benefit analysis and the new textualism. This Article contends with this problem by exploiting a unique event in administrative law: the Supreme Court's 2011 decision in Mayo Foundation v. United States, which required that courts apply Chevron deference to interpretative tax regulations. By altering the deference regime applicable to one specific category of regulation, Mayo created a natural experiment with a treatment group (interpretative tax regulations) and a control group (all other regulations).

This Article uses natural language processing and various statistical methods to evaluate the causal effect of Chevron deference. These techniques allow the Article to analyze a dataset of 69,956 regulations in a transparent and replicable manner. The Article finds that, after Mayo, the Department of the Treasury shifted its explanations for new tax rules to focus more on normative policy concerns and less on statutory interpretation. These results are statistically significant and large in magnitude: a 137.3% increase in language discussing

[†] Associate Professor, University of Minnesota Law School. Thanks to Edith Beerdsen, Tommy Bennett, Bryan Camp, Ryan Fackler, Andy Grewal, Kristin Hickman, Leandra Lederman, Josh Macey, Leigh Osofsky, Gil Rothenberg, Holger Spamann, Clint Wallace, Chris Walker, Eleanor Wilking, David Zaring, and the participants in the New York University School of Law Lawyering Scholarship Colloquium, the Florida-Michigan-Virginia Virtual Law and Economics Workshop, the University of Minnesota Law School Faculty Squaretable, the University of Oregon Law Tax Policy Colloquium, the Association of American Law Schools New Voices in Administrative Law and Legislation and New Voices in Taxation panels, the National Tax Association annual conference, and the Duke University Annual Administrative Law Symposium, for their useful comments on this paper. Last but not least, thanks to the editors of the *Yale Journal on Regulation* for excellent comments and editorial work.

normative goals (like efficiency and fairness) and a 48.5% decrease in language discussing the underlying statute.

In addition, the Article introduces a new theoretical model in which greater judicial deference encourages agencies to exert more effort in following rulemaking procedures. It hypothesizes that agencies counterintuitively view greater procedural effort as the “price” of judicial deference, a price that is more worth paying when courts are more deferential. Empirical analysis supports this hypothesis, finding that the move to Chevron deference caused a 18.9% increase in the length of regulatory preambles and a 35.8% increase in the intensity of preambles’ discussion of public comments.

These results cast new light on the debate over Chevron, suggesting that Chevron makes agency rulemaking more detailed and policy-focused. This raises the stakes of a potential Chevron reversal and clarifies the arguments of its supporters and critics. Ultimately, this Article underscores the importance of judicial deference regimes in shaping agency behavior.

Introduction.....	820
I. <i>Chevron</i> , <i>National Muffler</i> , and Rulemaking Style.....	824
A. Rulemaking as Policymaking Rather than Legal Interpretation	824
B. Procedural Effort as the Price of Deference	828
C. Does <i>Chevron</i> Matter?	834
D. <i>National Muffler</i> and Interpretative Tax Regulations	836
II. Data and Methods	839
A. Quantifying Text.....	839
B. Preambles as Windows into the Regulatory Process	842
C. Difference in Differences	843
D. Potential Confounders	846
E. Limitations	847
III. Results	849
A. Statutory and Normative Terms	849
B. Procedural Effort	852
C. Case Studies.....	855
IV. Specification Checks	857
A. Reading Preambles to Confirm Proxy Terms.....	857
B. Regulatory Selection Effects.....	857
C. Legislative History as an Alternative Proxy for Statutory Terms	859
D. Shortened-Window Regression Analysis, 2006-2015.....	860
E. Winsorized Regression Analysis	862
F. Placebo Tests.....	863
G. Extrapolating Pre-Trends	864
H. Synthetic Controls	865
Conclusion	869
Appendix	871

A. Modeling the Price of Deference.....	871
B. Data	873
C. Terms Analyzed	873
1. Normative Terms.....	874
2. Statutory Terms.....	874
3. Legislative History	875
4. Commenters.....	876
D. Distinguishing Interpretative and Legislative Regulations.....	876
E. Descriptive Statistics.....	880
F. Non-Normal Distribution of Term Frequencies	882
G. Regression Models: OLS, GLM, and Two-Part Regression.....	884
H. Choice of GLM Link Function and Distribution Family	888
I. Regression Tables.....	890

Introduction

*Chevron*¹ deference is under attack. Although *Chevron* is the most-cited administrative law case of all time² and one of the most important Supreme Court cases of any kind,³ judges,⁴ legislators,⁵ and scholars.⁶ over the past decade have increasingly called for its reversal. The appointments of *Chevron*-skeptical

1. *Chevron, U.S.A., Inc. v. Natural Resources Defense Council, Inc.*, 467 U.S. 837 (1984).

2. Peter M. Shane & Christopher J. Walker, *Foreword: Chevron at 30: Looking Back and Looking Forward*, 83 FORDHAM L. REV. 475, 475 (2014).

3. Thomas J. Miles & Cass R. Sunstein, *Do Judges Make Regulatory Policy? An Empirical Investigation of Chevron*, 73 U. CHI. L. REV. 823, 823 (2006) (describing *Chevron* as “the most-cited case in modern public law”).

4. *See, e.g.*, *Pereira v. Sessions*, 138 S. Ct. 2105, 2121 (2018) (“[I]t seems necessary and appropriate to reconsider, in an appropriate case, the premises that underlie *Chevron* and how courts have implemented that decision.”) (Kennedy, J., concurring); *id.* at 2129 (Alito, J., dissenting) (“In recent years, several Members of this Court have questioned *Chevron*’s foundations.”); *Michigan v. EPA*, 135 S. Ct. 2699, 2712 (2015) (Thomas, J., concurring) (“*Chevron* deference raises serious separation-of-powers questions.”); *Gutierrez-Brizuela v. Lynch*, 834 F.3d 1142, 1152 (10th Cir. 2016) (Gorsuch, J., concurring) (describing *Chevron* as “a judge-made doctrine for the abdication of the judicial duty”); Brett M. Kavanaugh, Book Review, *Fixing Statutory Interpretation*, 129 HARV. L. REV. 2118, 2154 (2016) (criticizing *Chevron* as “indeterminate” and “antithetical to the neutral, impartial rule of law”).

5. Separation of Powers Restoration Act of 2016, S. 2724, 114th Cong. § 2 (2016) (proposing to repeal deference by requiring courts to “decide de novo all relevant questions of law, including the interpretation of constitutional and statutory provisions, and rules made by agencies”); Separation of Powers Restoration Act of 2016, H.R. 4768, 114th Cong. § 2 (2016) (same).

6. *See, e.g.*, PHILIP HAMBURGER, IS ADMINISTRATIVE LAW UNLAWFUL? 316 (2014) (“[T]he deference to [administrative] interpretation is an abandonment of judicial office”); Jack M. Beermann, *End the Failed Chevron Experiment Now: How Chevron Has Failed and Why It Can and Should Be Overruled*, 42 CONN. L. REV. 779 (2010). Many other scholars have observed, but not necessarily encouraged, the decline of *Chevron*. Michael Herz, *Chevron Is Dead; Long Live Chevron*, 115 COLUM. L. REV. 1867, 1868 (2015) (“[R]eports of *Chevron*’s death seemed to get significant confirmation at the end of the Supreme Court’s 2014–2015 Term”); Cass R. Sunstein, *Chevron Without Chevron*, 2018 SUP. CT. REV. 59, 60 (“It seems clear that *Chevron* is entering a period of serious reconsideration. In the fullness of time, it might be seriously qualified or even abandoned.”); Linda Jellum, *Chevron’s Demise: A Survey of Chevron from Infancy to Senescence*, 59 ADMIN. L. REV. 725, 727 (2007) (describing “*Chevron*’s demise”).

Justices Gorsuch and Kavanaugh to the Supreme Court have inspired fresh arguments about judicial deference in general and *Chevron* deference in particular.

But despite the enormous volume of scholarly literature on *Chevron*,⁷ almost no empirical work has studied the effect of *Chevron* on agencies themselves. The only study so far to address this question—Christopher Walker’s survey of agency rule drafters—was inconclusive,⁸ leaving judges and policymakers in the dark about *Chevron*’s actual effects.

One common critique of *Chevron* is that it encourages “executive bureaucracies to swallow huge amounts of core judicial and legislative power.”⁹ More specifically, critics like Justice Thomas argue that it empowers agencies “not to find the best meaning of the text, but to formulate legally binding rules to fill in gaps based on policy judgments made by the agency rather than Congress.”¹⁰ This view is shared by many scholars, both critics and supporters of *Chevron*.¹¹ They believe that *Chevron* dramatically transformed agency rulemaking, from the province of lawyers seeking the most accurate reading of the statute into the province of technocratic agency experts seeking the normatively best policy.¹² Yet this view has been supported so far primarily by anecdote rather than empirical evidence.¹³

Another area that has received relatively little attention has been the relationship between *Chevron* deference and the requirements of procedural rulemaking, particularly agencies’ responsibilities to conduct notice and comment and adequately explain new regulations to the public. These procedures play a key role in administrative law—they are thought to increase public engagement, democratic accountability, agency legitimacy, the diversity of views considered in

7. Jerry Mashaw once joked that scholars have spilled so much ink over *Chevron* that it should have been issued with its own environmental impact statement. JERRY MASHAW, GREED, CHAOS, AND GOVERNANCE: USING PUBLIC CHOICE TO IMPROVE PUBLIC LAW 222 n.19 (1997).

8. *Infra* notes 42-46 and accompanying text.

9. *Gutierrez-Brizuela*, 834 F.3d at 1149-58 (Gorsuch, J., concurring) (“*Chevron* and *Brand X* permit executive bureaucracies to swallow huge amounts of core judicial and legislative power and concentrate federal power in a way that seems more than a little difficult to square with the Constitution . . .”).

10. *Michigan*, 135 S. Ct. at 2713 (Thomas, J., concurring).

11. *Infra* Section I.A.

12. E.g., E. Donald Elliott, *Chevron Matters: How the Chevron Doctrine Redefined the Roles of Congress, Courts and Agencies in Environmental Law*, 16 VILL. ENVTL. L.J. 1, 12 (2005) (“*Chevron* opened up and validated a policy making dialogue within agencies about what interpretation the agency should adopt for policy reasons, rather than what interpretation the agency must adopt for legal reasons. . . . *Chevron* has increased the weight given to the views of air pollution experts in the air program office relative to the lawyers in OGC.”).

13. Kavanaugh, *supra* note 4, at 2150 (“From my more than five years of experience at the White House, I can confidently say that *Chevron* encourages the Executive Branch (whichever party controls it) to be extremely aggressive in seeking to squeeze its policy goals into ill-fitting statutory authorizations and restraints.”); David S. Tatel, *The Administrative Process and the Rule of Environmental Law*, 34 HARV. ENVTL. L. REV. 1, 2 (2010) (“[I]t looks for all the world like agencies choose their policy first and then later seek to defend its legality.”); Elliott, *supra* note 12, at 11-12.

the regulatory process, and the quality of the regulation ultimately produced.¹⁴ But scholars have not yet studied *Chevron's* role in this process. In theory, *Chevron* is a doctrine of statutory interpretation entirely separate from rulemaking procedures, which are dictated by *State Farm*¹⁵ and section 553(c) of the Administrative Procedure Act (APA).¹⁶ *Chevron* therefore might have no effect on agency rulemaking efforts, or it might even discourage agencies from detailed compliance by making it less likely that courts will scrutinize agency rulemaking at all.

This Article argues that, counterintuitively, *Chevron* should encourage agencies to exert *more* effort in complying with rulemaking procedures, rather than less. This is because agencies will view procedural effort as essentially the price of judicial deference—even if *Chevron* and rulemaking requirements are theoretically separate, investment in procedural compliance is more worthwhile if the resulting regulation will receive more deferential review. This Article presents a new theoretical model that supports this hypothesis.¹⁷ It also argues that recent Supreme Court jurisprudence has further strengthened the link between *Chevron* deference and notice and comment requirements.¹⁸

On the other hand, many scholars and judges have argued that *Chevron's* influence is overstated. Some have alleged that *Chevron* is no more deferential than its alternatives: Justice Breyer and former Judge Posner have argued that *Chevron* deference is similar to *Skidmore* deference,¹⁹ while others have argued that *Chevron's* analysis of reasonableness is indistinguishable from the “arbitrary and capricious” analysis required under the APA, independent of *Chevron*.²⁰ Moreover, many scholars have empirically found that courts invoke *Chevron* less often than is commonly supposed, and that its influence is minimal even when invoked.²¹ All these perspectives suggest that *Chevron* might have little or no effect—that it might inspire no shift toward a policy focus and might inspire no additional procedural effort by agencies.

The essential problem in conducting causal analysis of *Chevron* has been the absence of a control group. Because, in theory, *Chevron* applied to all agencies at once, we cannot see how agency rulemaking would have developed counterfactually. This has made it difficult to separate the effect of *Chevron*

14. E.g., Michael Abramowicz & Thomas B. Colby, *Notice-and-Comment Judicial Decisionmaking*, 76 U. CHI. L. REV. 965, 968-69 (2009) (“In administrative law, the notice-and-comment process serves several related functions: providing information to decisionmakers, legitimating the decisionmaking process, and constraining decisionmakers by pushing them to confront arguments that point away from their preferred course of action.”); Gillian E. Metzger, *Through the Looking Glass to a Shared Reflection: The Evolving Relationship Between Administrative Law and Financial Regulation*, 78 L. & CONTEMP. PROBS. 129, 130 (2015) (describing the belief that notice and comment enhances political accountability of agencies as an “obsession” among administrative law scholars).

15. *Motor Vehicles Mfrs. Ass’n v. State Farm Mut. Auto. Ins. Co.*, 463 U.S. 29 (1983).

16. 5 U.S.C. § 553(c) (2018).

17. *Infra* Section I.B; Appendix Section A.

18. *Infra* notes 56-60 and accompanying text.

19. *Infra* note 78 and accompanying text.

20. *Infra* notes 79-80 and accompanying text.

21. *Infra* Section I.C.

specifically from contemporaneous events in the mid-1980s, like the rise of cost-benefit analysis,²² law and economics,²³ and the new textualism.²⁴

To address this problem, this Article proposes novel treatment and control groups: the IRS and other federal agencies. Courts have historically accorded a unique, lower degree of deference (*National Muffler* deference²⁵) to “interpretative tax regulations” promulgated under Treasury’s general power to make tax regulations.²⁶ This changed in 2011, when the Supreme Court ruled in *Mayo Foundation v. United States*²⁷ that all tax regulations are subject to *Chevron* deference, including interpretative regulations.²⁸ *Mayo* marked a huge shift in administrative tax law, but notably it marked a shift *only* in tax law, since regulations issued by other agencies had long been accorded *Chevron* deference.

This Article assembles a dataset comprising all the regulations digitally available on FederalRegister.Gov, an official governmental source of federal regulations from 2000 to the present. It algorithmically analyzes 69,956 discrete agency rules during that period, issued by every federal agency, quantifying various aspects of the regulatory preambles used to explain new regulations. It evaluates the relative importance of legal analysis and policy analysis by analyzing how frequently agencies discuss the relevant statute (reflecting a legal orientation) versus normative considerations like fairness and efficiency (reflecting a policy orientation). And it measures the effect of *Chevron* on agencies’ procedural effort by examining changes to the length of preambles and the frequency with which preambles discuss public comments. By quantifying text in this way, the Article facilitates both visual examination of trends and more complex statistical analysis.

The Article finds some evidence that the shift to *Chevron* deference caused the IRS to become more focused on policy issues and less focused on legal issues, and to exert greater procedural effort. Estimates of these effects are large in magnitude and statistically significant: a 137.3% increase in the frequency of normative terms, a 48.5% decrease in the frequency of statutory terms, a 18.9% increase in preamble length, and a 35.8% increase in the intensity of discussion of public comments.

Moreover, the Article investigates these results along multiple dimensions. *Mayo* made preambles to interpretative tax regulations more likely to use any normative terms at all and less likely to use any statutory terms at all. In addition, among interpretative tax preambles that discussed normative concepts (or statutory concepts), *Mayo* increased the *intensity* of use by increasing the

22. Exec. Order No. 12,291, 46 Fed. Reg. 13,193 (Feb. 17, 1981).

23. *E.g.*, RICHARD A. POSNER, *THE ECONOMICS OF JUSTICE* (1983).

24. Most prominently, Justice Scalia was appointed in 1986 and quickly became a prominent new textualist on the modern Court.

25. *Nat’l Muffler Dealers Ass’n, Inc. v. United States*, 440 U.S. 472 (1979).

26. *Infra* Section I.D.

27. 562 U.S. 44 (2011).

28. *Infra* Section I.D.

frequency of normative terms (and decreasing the frequency of statutory terms) within each preamble. In contrast, *Mayo* did not significantly increase the likelihood that a preamble referred to public comments at all; but among preambles referring to public comments, *Mayo* substantially increased the frequency of those references. The specific form of statistical analysis used in this Article therefore allows more sophisticated analysis of *Mayo*'s effects.

These results suggest that *Chevron* has a substantial impact on agency rulemaking. They sharpen the positions of *Chevron*'s critics and supporters, underscoring the importance of *Chevron* and raising the stakes of the current debate over judicial deference. They also add to the literature on tax exceptionalism by illustrating how tax law's unique deference regime prior to *Mayo* influenced Treasury's approach to drafting regulations.

This Article makes three main contributions to the existing literature on the effects of *Chevron*. First, it empirically supports claims that *Chevron* pushes agencies away from legal questions and toward policy ones. Second, it proposes a new theoretical model under which *Chevron* deference increases agencies' compliance with rulemaking procedures, because those procedures serve as the price of heightened deference. Third, it empirically supports the price-of-deference model, finding that *Chevron* deference does in fact increase the length of regulatory preambles and the extent to which agencies engage with public comments.

Part II discusses the history of and scholarly views on judicial deference and introduces the price-of-deference model. Part III describes the empirical methods used in this Article, and Part IV presents empirical results. Part V describes specification checks that largely support the results in Part IV but that also raise some notes of caution, especially with respect to the analysis of preamble length. Part VI concludes by discussing the implications of this Article's findings. The Appendix provides additional detail on mathematical proofs, data, and methods.

I. *Chevron*, *National Muffler*, and Rulemaking Style

A. *Rulemaking as Policymaking Rather than Legal Interpretation*

When an administrative agency issues regulations, does it make a legal judgment or a policy judgment?²⁹ Does it look to evidence of statutory meaning—

29. Cf. Catherine M. Sharkey, *Cutting in on the Chevron Two-Step*, 86 FORDHAM L. REV. 2359, 2365 (2018) (contrasting “legal” and “policy” rationales for rulemaking); Mark Seidenfeld, *A Syncopated Chevron: Emphasizing Reasoned Decision-Making in Reviewing Agency Interpretations of Statutes*, 73 TEX. L. REV. 83, 88-90 (1994) (phrasing the choice as between a “formalist or transmission-belt model” under which “administrative agencies expediently implement the will of the legislature” and an “expertise model” which “emphasizes the experience and technical knowledge of agencies and their staffs”).

statutory text and purpose, the intent of Congress³⁰—or does it ask which rule is normatively best, regardless of the statute?³¹

Chevron seemed to nudge agencies away from a statutory, legal orientation, and toward a normative, policy one. The Court held that courts must defer to an agency rule so long as the underlying statute is “silent or ambiguous” and the rule reflects a “reasonable policy choice.”³² The Court’s decision in *Chevron* explicitly noted that “an agency to which Congress has delegated policy-making responsibilities may, within the limits of that delegation, properly rely upon the incumbent administration’s views of wise policy to inform its judgments.”³³

Donald Elliott, a former Environmental Protection Agency (EPA) lawyer, recounts that prior to *Chevron*, the EPA had treated each statute as a “prescriptive text having a single meaning, discoverable by specialized legal training and tools.”³⁴ After *Chevron*, the EPA treated statutes as creating “a range of permissible interpretive discretion,” within which “[t]he agency’s policy-makers, not its lawyers, should decide which of several different but legally defensible interpretations to adopt.”³⁵

In the statutory interpretation literature, the pre-*Chevron* approach is sometimes described as the “faithful agent” model, where interpreters do not shape the law according to their own preferences, but instead try to extract meaning from statutes as accurately as possible (using any interpretive theory, whether textualist, purposivist, or otherwise).³⁶ Elliott suggested that agencies turned away from this model after *Chevron*, substituting their own value judgments for those of Congress. While we should not overstate the starkness of the shift—agencies will generally consider both policy and legal issues in their

30. Textualists, purposivists, and intentionalists debate the extent to which each of these should be used as evidence of statutory meaning. However, each of these schools of thought is ultimately interested in determining the objective meaning of the statute, rather than in de novo policymaking.

31. Of course, agencies ordinarily do both to some extent. As described further below in this Section, under *Chevron*, agencies may exercise normative discretion to pick among a range of options within a “*Chevron* space” determined through traditional tools of statutory interpretation. See *infra* notes 37-38 and accompanying text. But outside that *Chevron* space, normal tools of statutory interpretation will still apply. This Article studies the *balance* between legal analysis and policy analysis, recognizing that no agency will ever pursue only one or the other.

32. *Chevron, U.S.A., Inc. v. Nat. Res. Def. Council, Inc.*, 467 U.S. 837, 842, 843 & n.9 (1984).

33. *Id.* at 865-66.

34. Elliott, *supra* note 12, at 11.

35. *Id.* at 12; see also Elizabeth Magill & Adrian Vermeule, *Allocating Power Within Agencies*, 120 YALE L.J. 1032, 1046 (2011) (“As compared to the predecessor regime, a major effect of *Chevron* is to disempower lawyers within agencies.”); Jerry L. Mashaw, *Between Facts and Norms: Agency Statutory Interpretation as an Autonomous Enterprise*, 55 U. TORONTO L.J. 497, 532-33 & nn.71, 73 (2005). The dichotomy between lawyers and subject-matter experts becomes blurred at the IRS, where lawyers make *both* policy decisions *and* legal decisions. However, even at the IRS, there is a distinction between rulemaking on policy grounds and on legal grounds.

36. Cass R. Sunstein, *Interpreting Statutes in the Regulatory State*, 103 HARV. L. REV. 405, 415 (1989); John F. Manning, *The New Purposivism*, 2011 SUP. CT. REV. 113, 119-20; Jonathan H. Choi, *An Empirical Study of Statutory Interpretation in Tax Law*, 95 N.Y.U. L. REV. 363, 368 n.15 (2020).

rulemaking—Elliott’s view, common among administrative scholars, is that *Chevron* initiated a significant move away from law and toward policy.

Peter Strauss similarly describes judicial deference as creating a “*Chevron* space.” In his view, readings of the statute that are “permissible” but not “necessary” fall within this *Chevron* space, and agencies may freely select among these readings without judicial override.³⁷ This provides agencies the opportunity to make rules on policy grounds, if they prefer. In the same vein, other scholars have created theoretical models that assume agencies optimize policy goals rather than attempting to interpret statutes as accurately as possible.³⁸

Opponents of *Chevron* often criticize agencies for replacing congressional policy judgments with agencies’ own.³⁹ In contrast, supporters of *Chevron* often celebrate it on the same basis—that agencies can bring field-specific expertise that Congress and judges lack, and that considering policy goals in rulemaking will help to implement statutes as effectively as possible.⁴⁰

But although many claim that *Chevron* has inspired a shift from legal to policy analysis, there are some dissenting voices and substantial reason to doubt the extent of *Chevron*’s influence. Although most would agree that *Chevron* de-emphasizes legal analysis of statutes, some would disagree that it elevates normative policy analysis instead. Catherine Sharkey argues that because *Chevron* step two conventionally does not require courts to assess the normative validity of agency rulemaking, agencies have little incentive to explain themselves or to make rules on appropriate policy grounds.⁴¹ Within the *Chevron* space, it could be that agencies make rules for less savory reasons (self-interest, political expediency) or for no apparent reason.

37. Peter L. Strauss, “Deference” Is Too Confusing—Let’s Call Them “Chevron Space” and “Skidmore Weight,” 112 COLUM. L. REV. 1143, 1163-64 (2013).

38. Matthew C. Stephenson, The Strategic Substitution Effect: Textual Plausibility, Procedural Formality, and Judicial Review of Agency Statutory Interpretations, 120 HARV. L. REV. 528, 535, 536, 544 (2006) (assuming that agencies are “interpretive instrumentalists, attaching no intrinsic importance to textual fidelity or analogous concerns” but instead attempting to “secure whatever interpretation would best advance its substantive policy agenda”); John R. Wright, Ambiguous Statutes and Judicial Deference to Federal Agencies, 22 J. THEORETICAL POL. 217, 226 (2010) (also modelling agency action as a function of policy goals).

39. See *supra* note 13 and accompanying text; see also *Michigan v. EPA*, 135 S. Ct. 2699, 2713 (2015) (Thomas, J., concurring) (complaining that *Chevron* empowers agencies “not to find the best meaning of the text, but to formulate legally binding rules to fill in gaps based on policy judgments made by the agency rather than Congress”); Beermann, *supra* note 6, at 784 (“*Chevron* encourages irresponsible agency and judicial behavior. Agencies expecting that their interpretive decisions will be reviewed under a deferential version of *Chevron* are free to disregard congressional intent and impose their own policy views even when it is possible to have at least a good sense of how Congress would have wanted the agency to act.”).

40. Cass R. Sunstein & Adrian Vermeule, *The Unbearable Rightness of Auer*, 84 U. CHI. L. REV. 297, 305 (2017) (“When a statute is unclear, and especially when a complex modern regulatory statute is unclear, resolution of the ambiguity will inevitably require policy-making competence—which courts lack and which agencies have.”); Cass R. Sunstein & Adrian Vermeule, *Interpretation and Institutions*, 101 MICH. L. REV. 885, 928-30 (2003) (describing, with approval, an EPA regulation formulated on pragmatic public policy rather than purely statutory grounds).

41. Sharkey, *supra* note 29, at 2370-73.

Those who have described a shift toward normative rulemaking have primarily relied on anecdote.⁴² While empirical literature suggests that both congressional drafters and agency administrators are aware of *Chevron*,⁴³ past work has been agnostic regarding *Chevron*'s actual effect on agency activity. Christopher Walker's 2015 survey of agency rule drafters touches on this question but produced mixed results. Walker asked agency administrators whether heightened judicial deference would make an agency "more willing to advance a more aggressive interpretation."⁴⁴ Willingness to advance a more aggressive interpretation is a necessary but not sufficient condition for an agency to shift its focus to policy. For example, administrators might adopt an aggressive interpretation not to enact ideal policies, but rather to entrench their own power, or advance other selfish career goals at the expense of public administration.⁴⁵ Walker's survey results were inconclusive; while two in five administrators agreed with this statement, others indicated that:

the judicial deference standard is just one of many factors that affect agency statutory interpretation, and it may be a pretty insignificant factor in the large scheme. And a couple rule drafters indicated that they had never personally taken into account or observed others taking into account the type of deference the agency expected to receive.⁴⁶

Walker concludes that "broader generalizations about whether agencies draft more aggressively when they know *Chevron* applies probably cannot be drawn from this study," suggesting that "the findings uncovered should encourage deeper empirical inquiry."⁴⁷

A prior article of mine was one of the earliest efforts to empirically study agencies' relative focus on legal interpretation and policymaking, based on regulations themselves.⁴⁸ It revealed that the IRS became substantially more normative and less statutory in the guidance that it issued during the 1980s.⁴⁹

42. Kavanaugh, *supra* note 4; Elliott, *supra* note 12.

43. Lisa Schultz Bressman & Abbe R. Gluck, Statutory Interpretation from the Inside—An Empirical Study of Congressional Drafting, Delegation and the Canons: Part I, 66 STAN. L. REV. 901, 927 (2013); Christopher J. Walker, Inside Agency Statutory Interpretation, 67 STAN. L. REV. 999, 1019 (2015).

44. Walker, *supra* note 43, at 1063; Christopher J Walker, *Chevron Inside the Regulatory State: An Empirical Assessment*, 83 FORDHAM L. REV. 703, 723-24 (2014).

45. Walker's hypothesis that agency interpretations might be more "aggressive" when judicial deference is stronger follows a long line of political science literature that generally models the relationship between Congress and agencies as one between principals and agents, each of which is "assumed to have different preferences." RACHEL AUGUSTINE POTTER, BENDING THE RULES: PROCEDURAL POLITICKING IN THE BUREAUCRACY 7 (2019) (describing an "external perspective" in the political science literature, "premised on a principal-agent framework").

46. Christopher J Walker, *Chevron Inside the Regulatory State: An Empirical Assessment*, 83 FORDHAM L. REV. 703, 724 (2014).

47. *Id.* at 725.

48. Choi, *supra* note 36.

49. *Id.* at 392-95.

However, this study had several features that prevented it from serving as a study of the causal effect of *Chevron* on agency rulemaking.

First, because it only considered tax regulations, it failed to compare trends in IRS guidance with trends at other agencies. Second, because it examined only broad trends over time, there was no way to separate the effect of *Chevron* from other important contemporaneous developments that could have affected IRS guidance, like the rise of cost-benefit analysis, the appointment of new textualist judges, and the popularization of law and economics.⁵⁰ As a result, this study did not test the hypothesis that *Chevron* itself actually caused a normative shift in the 1980s. Third, as Section I.D discusses, most Treasury regulations were not thought to be subject to *Chevron* deference at all prior to 2011. So, this prior study not only refrained from claims of causal inference regarding the effect of judicial deference in general; it also refrained from claims of causal inference relating to *Chevron* deference in particular.⁵¹

Consequently, research discussing *Chevron*'s effect on the balance between normative policymaking and statutory interpretation remains incomplete. While *Chevron* is sometimes believed to permit agencies to make rules based on policy considerations rather than legal ones, theoretical accounts are mixed and empirical evidence on this question is thin.

B. Procedural Effort as the Price of Deference

Another way that *Chevron* might affect agency rulemaking is by changing the amount of effort that agencies exert in complying with rulemaking procedures, especially when they explain new regulations and engage with public comments.⁵²

The simplest hypothesis is that *Chevron* does not matter. The obligations to adequately explain rulemaking and address public comments are procedural requirements under *State Farm*⁵³ and section 553(c) of the APA.⁵⁴ In theory, they are independent obligations that would apply equally under *Chevron* or any other deference regime.

Another view is that *Chevron* provides a pure benefit—that it is a boon to agencies, providing them additional cover to advance their own regulatory preferences. On this theory, a shift to *Chevron* deference might embolden agencies to write shorter preambles that engage less with public comments. Writing a preamble is time-consuming, and reading voluminous public comments

50. *Id.* at 393.

51. *Id.* at 393-95.

52. Procedural effort contrasts with what one might call “substantive effort.” The former focuses on the procedural aspects of regulations, especially the process of explaining and justifying regulations through preambles; the latter would focus on the drafting of regulations themselves. Procedural effort may or may not serve as a proxy for rulemaking effort in general.

53. *Motor Vehicles Mfrs. Ass'n v. State Farm Mut. Auto. Ins. Co.*, 463 U.S. 29 (1983).

54. Section 553(c) of the APA requires that an agency adopting a rule through notice and comment provide a “concise general statement of [the rule's] basis and purpose.” 5 U.S.C. § 553(c) (2018).

and substantively responding to them can be a daunting task for agency employees.⁵⁵ A lengthy and responsive preamble might be necessary if *Chevron* does not apply and courts are likely to scrutinize the rationale behind agency rulemaking. But why bother writing long, technical preambles if your regulation will be upheld regardless?

This Article argues the opposite—that we should expect heightened judicial deference to *increase* the amount of effort agencies exert in procedural compliance. The Supreme Court has moved in recent decades to strengthen the link between procedural rulemaking requirements and *Chevron*. Most notably, the Supreme Court has linked *Chevron* deference with the notice and comment process. The Court said in 2001 that notice and comment is a “very good indicator” that a regulation is intended to have the force of law and therefore receive *Chevron* deference.⁵⁶

Similarly, the Court ruled in 2016 that “*Chevron* deference is not warranted where the regulation is ‘procedurally defective’—that is, where the agency errs by failing to follow the correct procedures in issuing the regulation.”⁵⁷ It went on to say that a regulation can be procedurally defective, forfeiting *Chevron* deference, if the agency fails to “give adequate reasons for its decisions.”⁵⁸ Courts also sometimes merge *State Farm* analysis with *Chevron* step two,⁵⁹ and the Supreme Court’s considerable jurisprudence surrounding “*Chevron* step zero”—the initial judicial determination over whether *Chevron* should apply—has suggested that

55. POTTER, *supra* note 45, at 33.

56. United States v. Mead Corp., 533 U.S. 218, 229-30 (2001); *see also* Encino Motorcars, LLC v. Navarro, 136 S. Ct. 2117, 2125 (2016) (citing *Mead*). Relatedly, several scholars have proposed that regulations should receive *Chevron* deference if and only if they are issued with notice and comment. This proposal, advanced by scholars including Jacob Gersen and John Manning, has been dubbed the “short cut.” *See* David L. Franklin, *Legislative Rules, Nonlegislative Rules, and the Perils of the Short Cut*, 120 YALE L.J. 276, 299 (2010) (coining the term “short cut”); Jacob E. Gersen, *Legislative Rules Revisited*, 74 U. CHI. L. REV. 1705, 1719 (2007); John F. Manning, *Nonlegislative Rules*, 72 GEO. WASH. L. REV. 893, 931 (2004).

The “short cut” is distinct from the price-of-deference theory in that it is a normative proposal for what the law should be rather than a descriptive proposal for how agency incentives actually operate. Moreover, the short-cut theory primarily considers whether a regulation should be considered legislative and therefore receive *Chevron* deference. In contrast, the price-of-deference theory *assumes* that an agency does conduct notice and comment and that a given regulation will receive a fixed level of deference (*Chevron* or *National Muffler* in particular). Given this assumption, it then asks how much effort the agency will exert in the notice and comment process. Put differently, the key decision in the short-cut theory is whether or not to engage in notice and comment; the key decision in the price-of-deference theory is how much effort to exert in notice and comment, assuming that notice and comment occurs.

57. *Encino Motorcars*, 136 S. Ct. at 2125 (citing *Mead*, 533 U.S. at 227).

58. *Id.* Similarly, the Court ruled in 2015 that “[n]ot only must an agency’s decreed result be within the scope of its lawful authority, but the process by which it reaches that result must be logical and rational. It follows that agency action is lawful only if it rests ‘on a consideration of the relevant factors.’” *Michigan v. EPA*, 135 S. Ct. 2699, 2706 (2015) (quoting *Motor Vehicles Mfrs. Ass’n v. State Farm Mut. Auto. Ins. Co.*, 463 U.S. 29, 43 (1983)). This pair of cases has been taken by some scholars to suggest that the Court is reconsidering its *Chevron* jurisprudence to apply a more searching “hard look” review under *Chevron* step two. *E.g.*, Sharkey, *supra* note 29, at 2419-29. But the foundation for this move was laid much earlier, at the latest by *Mead* in 2001.

59. *E.g.*, Ronald M. Levin, *The Anatomy of Chevron: Step Two Reconsidered*, 72 CHI.-KENT. L. REV. 1253 (1997).

greater deliberation in rulemaking is a factor in applying *Chevron*.⁶⁰ Similarly, under the *Chenery* doctrine, courts reviewing agency rulemaking may only consider justifications for rules presented by the agencies themselves.⁶¹ Each of these judicial moves might link *Chevron* deference with rulemaking procedures in the minds of agencies.

Of course, we should not overinterpret particular phrases from a few key cases. The link between *Chevron* and rulemaking procedures remains ambiguous, and *Chevron* conventionally remains completely separate from *State Farm* and section 553(c) of the APA. Scholars attempting to link *Chevron* with rulemaking procedures generally frame their arguments as a new proposal or a path for future Supreme Court doctrine, rather than a claim about established doctrine.⁶²

This Article therefore advances a more ambitious claim: even assuming that *Chevron* imposes no procedural rulemaking requirements, greater interpretive deference through *Chevron* should result in greater procedural effort. This is because agencies will see longer and more responsive preambles as the price of deference: agencies should invest more resources into satisfying procedural requirements if the payoff is a regulation that will benefit from heightened deference. In this way, the price that agencies are willing to pay should increase if the deference regime becomes more favorable to them, much as consumers should be willing to pay a higher price for more useful goods.

It may seem counterintuitive that increased judicial deference could encourage an agency to spend greater effort to justify its rulemaking. The key insight is that under a weaker deference regime, agencies may not find it worthwhile to exert much effort obeying rulemaking procedures—procedural compliance might become irrelevant if the regulation is rejected on statutory interpretation grounds. Regardless, a higher degree of deference removes this risk, increasing the payoff from procedural effort.

To illustrate this point, Section A of the Appendix introduces a theoretical model of agency decision-making as a function of the amount of effort exerted in rulemaking procedures, the level of judicial deference, and the benefit from successfully promulgating a new regulation. The model includes two levels of judicial deference, *Chevron* and “sub-*Chevron*” (which could stand in for any less deferential regime, like *Skidmore*).⁶³ It is purely theoretical: it is loosely based on existing empirical scholarship⁶⁴ but does not draw on the empirical evidence in this Article.

60. Barnhart v. Walton, 535 U.S. 212 (2002); see also Michael Pollack & Daniel Hemel, *Chevron Step 0.5*, YALE J. ON REG.: NOTICE & COMMENT (June 24, 2016), <https://www.yalejreg.com/nc/chevron-step-0-5-by-michael-pollack-and-daniel-hemel> [<https://perma.cc/5G2T-EKKW>] (discussing *Encino Motorcars*).

61. SEC v. *Chenery Corp.*, 318 U.S. 80, 95 (1943). While *Chenery* predated the APA, it has subsequently been held to apply to formal agency rulemaking. See, e.g., *Vehicle Mfrs. Ass'n of the U.S. v. State Farm Mut. Auto. Ins. Co.*, 463 U.S. 29, 50 (1983); Kevin M. Stack, *The Constitutional Foundations of Chenery*, 116 YALE L.J. 952, 955-56 (2007) (describing the widespread application of *Chenery*).

62. E.g., Seidenfeld, *supra* note 29; Sharkey, *supra* note 29.

63. *Infra* app., eq. 2 and accompanying text.

64. See *infra* note 5 and accompanying text.

The model follows the standard view that *State Farm* and judicial interpretations of section 553(c) of the APA dictate agencies' obligation to explain rules and respond to comments, whereas judicial deference asks whether a regulation is substantively permissible in light of the statute. Consequently, the model takes the probability that a regulation will be upheld on judicial deference grounds as given (because it depends on factors outside the model) and treats the probability of compliance with *State Farm* and section 553(c) as a function of effort in explaining rules and responding to public comments. The overall probability that a regulation will be upheld is simply the product of these two probabilities.

Note that this model is agnostic about agency motivations. Agencies might desire good governance; or pursue self-interest (for example, they might aim to avoid the embarrassment of having regulations overturned); or view rulemaking as performative; or some combination of the three. The model merely assumes that agencies prefer for their regulations to be upheld.

The model has several features that mimic current regulatory practice. First, the probabilities have a minimum value above zero, reflecting a baseline likelihood that any particular regulation will be upheld, even with minimal agency effort. Second, the probability that rulemaking will be upheld under *State Farm* smoothly increases as a function of effort (in mathematical terms, it monotonically increases). Third, there are decreasing returns to additional procedural effort (in mathematical terms, the probability function is concave), so that the likelihood a regulation is upheld is ultimately bounded by a maximum likelihood depending on the deference regime (80% for *Chevron*, 50% for sub-*Chevron*).⁶⁵ Fourth, the likelihood that a regulation will be upheld is always higher where *Chevron* applies than under the less deferential regime. The model does *not* assume that *Chevron* is less deferential than the alternative at any point on the curve. Agencies might prefer to exert more effort even though, and indeed precisely because, *Chevron* is always more deferential at any given level of effort.

Given these probability functions, we can model agency action as an attempt to maximize utility as a function of procedural effort. Effort is directly costly, lowering utility; however, it also increases utility by increasing the likelihood that a regulation will be upheld.⁶⁶ Within this model, heightened judicial deference increases the optimal amount of effort that an agency exerts in rulemaking. Section A of the Appendix includes more formal mathematical discussion and a proof of this result. The proof depends on a limited set of assumptions, primarily monotonicity and concavity, and any probability function with these properties will satisfy the proof. Figure 1 graphically illustrates the probability that a

65. These percentages are loosely based on empirical findings by Kent Barnett and Chris Walker that agencies prevailed 77.4% of the time under *Chevron* and 56% of the time under *Skidmore* within their sample of cases. Kent Barnett & Christopher J. Walker, *Chevron in the Circuit Courts*, 116 MICH. L. REV. 1, 6 (2017). Because regulations challenged in court are likely the ones with the lowest probability of being upheld, the true percentage likelihood of prevailing is probably higher across the entire population of issued regulations.

66. See *infra* app., eq. 1 and accompanying text.

regulation will be upheld as a logistic function, and Figure 2 illustrates expected utility as a function of procedural effort, again using the logistic function. But any other concave, monotonically increasing function could also be used.

Figure 1: Probability that a Regulation Is Upheld as a Function of Procedural Effort

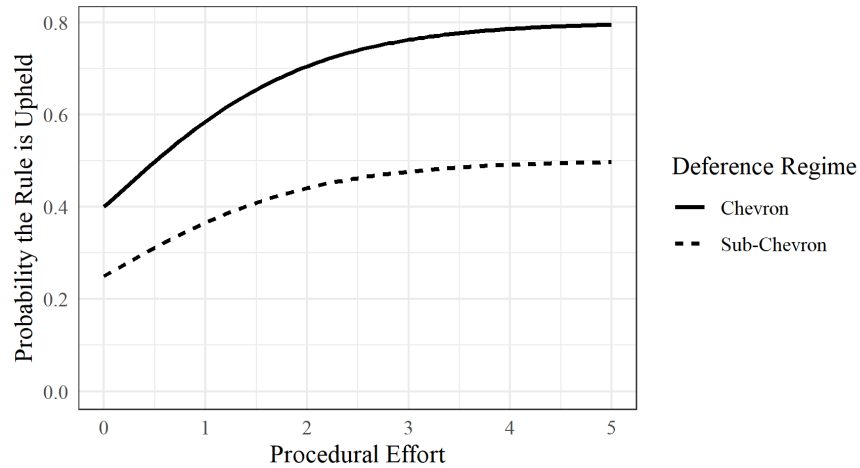
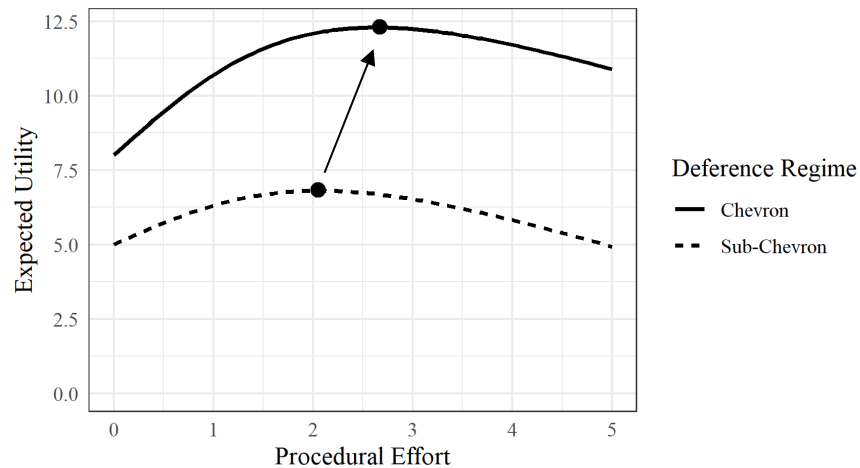


Figure 2: Expected Utility as a Function of Procedural Effort



The theory that procedural effort serves as the price of deference therefore operates along two potential paths. First, as the model suggests, even if judicial deference and rulemaking procedures are completely separate as a theoretical matter, agencies have an incentive to increase their procedural effort if judicial deference increases. Second, we should expect that this relationship has grown

even stronger in recent years, as the Supreme Court has gestured at a more explicit link between rulemaking procedures and *Chevron*.⁶⁷

Although regulatory procedures might seem secondary to the substantive content of regulations, procedures play an important role in the modern regulatory state. Oceans of scholarly ink have been spilled discussing the benefits of notice and comment: it facilitates public engagement and makes agencies democratically accountable, which scholars frequently cite as the primary justification for judicial deference to agency interpretation.⁶⁸ Public comments provide agencies with additional information and more diverse viewpoints, bolster agency legitimacy, and encourage agencies to engage in more deliberative policy analysis prior to issuing regulations.⁶⁹ The explanations that agencies offer in the rulemaking process are also themselves important interpretive references for lawyers and judges.⁷⁰

On the other hand, procedural effort has downsides as well. It is time-consuming and may not be an optimal use of agency resources. A considerable political science literature has debated whether procedural requirements like notice and comment ossify agency rulemaking by making it prohibitively expensive to repeal or amend existing regulations, or to propose new regulations.⁷¹ If judicial deference encourages agencies to spend even more time and resources on procedural aspects of rulemaking, it could potentially lead to even more ossification. Some scholars have also suggested that notice and comment may be a tool for agency capture because its participants are disproportionately powerful interest groups with access to high-priced lawyers.⁷² Still others have argued that regulatory preambles have become excessively long,

67. See *supra* notes 56-60 and accompanying text.

68. E.g., Metzger, *supra* note 15, at 130 (describing the belief that notice and comment enhances political accountability of agencies as an “obsession” among administrative law scholars); see also Lisa Schultz Bressman, *Procedures as Politics in Administrative Law*, 107 COLUM. L. REV. 1749, 1751 (2007) (describing some scholars’ belief that “administrative procedures may help to ensure that agencies stay more or less in line with legislative preferences”).

69. E.g., Abramowicz & Colby, *supra* note 15, at 968-69 (“In administrative law, the notice-and-comment process serves several related functions: providing information to decisionmakers, legitimating the decisionmaking process, and constraining decisionmakers by pushing them to confront arguments that point away from their preferred course of action.”).

70. Kevin M. Stack, *Interpreting Regulations*, 111 MICH. L. REV. 355, 361 (2012) (“[T]he text of a regulation and its statement of basis and purpose [preamble] stand in a unique relationship: together, they constitute the act of regulation, an act that is not complete without either element of this couplet. Based on this premise, it does not make sense to interpret the text of a regulation independently from its statement of basis and purpose.”); Kevin M Stack, *Preambles as Guidance*, 84 GEO. WASH. L. REV. 1252, 1252 (2016) (describing preambles as “the most authoritative source of guidance about the meaning of agency regulations”); see also Sharkey, *supra* note 29, at 2365 (describing how preambles serve as the basis for judicial oversight of regulations).

71. E.g., Jason Webb Yackee & Susan Webb Yackee, *Administrative Procedures and Bureaucratic Rulemaking: Is Federal Rule-Making “Ossified”?*, 20 J. PUB. ADMIN. RES. & THEORY 261 (2010).

72. Clinton G. Wallace, *Congressional Control of Tax Rulemaking*, 71 TAX L. REV. 179, 182 (2017) (finding, after reviewing a sample of tax regulations, that participation in notice and comment is “heavily weighted towards private interests”).

making them difficult to read and therefore less useful to lawyers and the general public.⁷³

Ultimately, as with the balance between statutory and normative analyses of rulemaking, opinions differ on the ideal amount of procedural effort that agencies should exert. But it is notable that prior literature has been so sparse on the link between judicial deference and rulemaking procedures. This Article introduces a new theoretical account within which such a link could plausibly exist, where more deference encourages more procedural effort.

Of course, this is purely a theoretical model, and heavily dependent on specific modelling choices.⁷⁴ In reality, judicial deference might have little or no effect on procedural effort, and the incentives described by the model could be counterbalanced by other considerations that the model does not address. Only empirical investigation can confirm whether heightened judicial deference actually encourages agencies to exert greater procedural effort.

C. Does Chevron Matter?

Empirical studies of *Chevron's* effect on *courts* have often argued that *Chevron's* influence is overstated, that courts invoke *Chevron* less often than is commonly assumed,⁷⁵ and that *Chevron* is rarely outcome determinative even when invoked.⁷⁶ This literature gives some reason to doubt that agencies really

73. E.g., Alec Webley, *Seeing Through a Preamble, Darkly: Administrative Verbosity in an Age of Populism and "Fake News,"* 70 ADMIN. L. REV. 1, 1 (2018) (“[T]he public truth-telling function of these ‘preambles’ has become undermined by their spectacular length, often to more than a thousand pages longer than their parent rules, making it virtually impossible for anyone (even lawyers!) to properly read them.”). On the other hand, preambles to tax regulations have historically erred on the side of brevity, nothing near the thousand-page preambles of other agencies.

74. See *infra* Appendix Section A.

75. William N. Eskridge & Lauren E. Baer, *The Continuum of Deference: Supreme Court Treatment of Agency Statutory Interpretations from Chevron to Hamdan*, 96 GEO. L.J. 1083, 1125 (2008) (“[T]he Court does not apply the Chevron framework in nearly three-quarters of the cases where it would appear applicable.”). But see Natalie Salmanowitz & Holger Spamann, *Does the Supreme Court Really Not Apply Chevron When It Should?*, 57 INT. REV. LAW ECON. 81, 81 (2019) (“Our reexamination of [Eskridge and Baer’s] study finds that the fraction of such cases is far lower, and indeed closer to zero.”).

76. E.g., Richard J. Pierce, Jr., *What Do the Studies of Judicial Review of Agency Actions Mean?*, 63 ADMIN. L. REV. 77, 93 (2011) (“There is no empirical support for the widespread belief that choice of doctrine plays a major role in judicial review of agency actions.”); David Zaring, *Reasonable Agencies*, 96 VA. L. REV. 135, 135 (2010) (“[T]he variance of the validation rates of agency action, regardless of the standard of review, is small.”); Ann Graham, *Searching for Chevron in Muddy Watters: The Roberts Court and Judicial Review of Agency Regulations*, 60 ADMIN. L. REV. 231, 271-72 (2008) (“*Chevron* has become the argument for the losing side, with failure by the majority to adhere to a straightforward *Chevron* analysis emerging as a recurring criticism in dissenting opinions.”); Richard W. Murphy, *Abandon Chevron and Modernize Stare Decisis for the Administrative State*, 69 ALA. L. REV. 1, 3 (2017) (“Notwithstanding overheated charges, there is little reason to think that applying *Chevron*, as opposed to a supposedly tighter standard of review, such as *Skidmore* deference, is frequently outcome determinative in significant cases.”); Eskridge & Baer, *supra* note 75, at 1142 (finding that the Supreme Court affirmed agency rulemaking at very similar rates whether applying *Chevron* or *Skidmore*). But see Barnett & Walker, *supra* note 65, at 6 (finding that different forms of judicial deference result in substantially different rates of affirmation at circuit courts); Kent Barnett, Christina L. Boyd & Christopher J. Walker, *Administrative Law’s Political Dynamics*, 71 VAND. L. REV. 1463, 1468 (2018) (“We find that *Chevron* deference significantly curbs (but does not fully constrain) judicial discretion.”).

alter their rulemaking practices in response to *Chevron*, if administrators believe that *Chevron* is ineffective (either based on their own observations or based on the empirical literature).

Separately, scholars and judges have provided additional theoretical reasons to doubt that *Chevron*'s imposition (or, conversely, its abolition) really changes the applicable level of judicial deference. Certain judges, including Justice Breyer and former Judge Posner, have suggested that *Chevron* is not in fact much more deferential than the standard one step down, *Skidmore*⁷⁷ deference.⁷⁸ And some scholars and courts have merged *Chevron*'s "reasonableness" standard⁷⁹ with the "arbitrary and capricious" standard that preceded it under the APA.⁸⁰ While *Chevron* remains frequently discussed and highly influential in spite of these critiques, they provide some reason to doubt that *Chevron* had the impact suggested by thinkers like Elliott⁸¹ or by the price-of-deference model.

Thus, a significant gap remains in the empirical literature on the effects of *Chevron*. Many theorists believe that *Chevron* encourages agencies to focus on policy issues rather than legal ones; but others disagree, and empirical evidence is limited on this point. I propose in this Article that *Chevron* encourages agencies to increase efforts in procedural aspects of rulemaking, but this new theory cuts against conventional wisdom and has not yet been tested. And, contrary to these

A related but separate line of scholarship considers the frequency of litigation challenges to agency rulemaking. See, e.g., Cary Coglianese & Daniel E. Walters, *Litigating EPA Rules: A Fifty-Year Retrospective of Environmental Rulemaking in the Courts*, 70 CASE W. L. REV. 1007 (2020) (empirically studying the rates at which EPA regulations were challenged in court and the frequency of success in these lawsuits).

77. *Skidmore v. Swift & Co.*, 323 U.S. 134, 140 (1944) (holding subregulatory guidance, "while not controlling upon the courts by reason of their authority, do[es] constitute a body of experience and informed judgment to which courts and litigants may properly resort for guidance"). Courts have long held that agency interpretations of ambiguous statutes might be "entitled to very great respect." *Edwards' Lessee v. Darby*, 25 U.S. (12 Wheat.) 206, 210 (1827); see also *Fawcus Mach. Co. v. United States*, 282 U.S. 375, 378 (1931) (holding that contemporaneous construction of an administering agency was "entitled to respectful consideration"); *Swendig v. Wash. Water Power Co.*, 265 U.S. 322, 331 (1924) (same).

78. *Murphy*, *supra* note 76, at 41-42.

79. E.g., *Judulang v. Holder*, 132 S. Ct. 476, 483 n.7 (2011) ("[U]nder *Chevron* step two, we ask whether an agency interpretation is 'arbitrary or capricious in substance.'"); *Chamber of Commerce v. FEC*, 76 F.3d 1234, 1235 (D.C. Cir. 1996) ("[T]he second step of *Chevron* . . . overlaps with the arbitrary and capricious standard"); Elizabeth V. Foote, *Statutory Interpretation or Public Administration: How Chevron Misconceives the Function of Agencies and Why It Matters*, 59 ADMIN. L. REV. 673, 710 (2007) (describing how certain scholars have argued that "arbitrary and capricious" is the appropriate standard for reasonableness under *Chevron* step two, and citing recent decisions by lower federal courts and the Supreme Court embracing this view); Ronald M. Levin, *The Anatomy of Chevron: Step Two Reconsidered*, 72 CHI.-KENT L. REV. 1253, 1254 (1997) (arguing that the "arbitrary and capricious" standard and reasonableness under *Chevron* "should be deemed not just overlapping, but identical").

80. Foote, *supra* note 79, at 675 ("Before *Chevron*, courts tended to use the statutory standard of arbitrary and capricious review and its close kin, the substantial evidence test, for oversight of most agency 'carrying out' actions—that is, for review of quintessential administrative implementation of statutory programs.").

81. Of course, it is possible that *Chevron* had an effect on agencies but not on courts, if agencies reformed their rulemaking mistakenly expecting to receive heightened judicial deference. Beermann, *supra* note 6, at 783 n.9. But this argument requires an additional inferential step and seems less plausible than *Chevron* directly awarding more deference to agency determinations.

theoretical predictions, much of the existing literature on *Chevron's* application by courts suggests that it may not matter at all.

D. National Muffler and Interpretative Tax Regulations

Administrative tax law presents a unique opportunity to measure the effects of a change in deference regime. Prior to 2011, the Treasury Department had long distinguished between “legislative” and “interpretative” regulations.⁸² Ostensibly, legislative regulations were those for which Congress delegated specific rulemaking authority to the IRS—for example, the grant in 26 U.S.C. § 1502 for the Secretary of the Treasury to “prescribe such regulations as he may deem necessary” to tax consolidated corporate groups. In contrast, interpretative rules did not require a specific act of rulemaking authority, but rather relied on the general instruction in 26 U.S.C. § 7805(a) that the Secretary “prescribe all needful rules and regulations for the enforcement of” the entire tax code.⁸³

The distinction in tax law between legislative rules based on specific congressional authorization and interpretative rules based only on general congressional authorization predated both the APA and *Chevron*.⁸⁴ Importantly, it differed from the mainstream definition of “interpretative” rules in administrative law, which defines regulations as legislative if they “carry the force of law.”⁸⁵ As Kristin Hickman argued prior to 2011, virtually all Treasury regulations (both interpretative and legislative) carried the force of law even prior to *Mayo*. Moreover, Treasury conducted notice and comment for virtually all Treasury regulations (both interpretative and legislative) both before and after *Mayo*,⁸⁶ another hallmark that a regulation is “legislative” in general

82. The terminology derives from the APA, which exempts “interpretative rules” from notice and comment requirements. 5 U.S.C. § 553(b)(A) (2018). Interpretative rules are sometimes also known as “interpretive” rules.

83. See, e.g., Bryan T. Camp, *A History of Tax Regulation Prior to the Administrative Procedure Act*, 63 DUKE L.J. 1673, 1673 (2014) (describing “the long-held view of the Department of Treasury that tax regulations issued under the general grant of authority in I.R.C. § 7805(a) are interpretative regulations within the meaning of the APA”); ABA Section of Taxation, *Report of the Task Force on Judicial Deference*, 57 TAX LAW. 717, 728 (2004) (“Interpretive regulations are those promulgated under the general authority of section 7805(a)”); Thomas W. Merrill & Kathryn Tongue Watts, *Agency Rules with the Force of Law: The Original Convention*, 116 HARV. L. REV. 467, 571-75 (2002).

84. Leandra Lederman, *The Fight Over “Fighting Regs” and Judicial Deference in Tax Litigation*, 92 B.U.L. REV. 643, 654-59 (2012). This distinction was unique to tax law. *Id.* at 656 (“Prior to the APA, there was an understanding specific to tax law that general-authority regulations were interpretative and that specific-authority regulations were legislative.”).

85. Ostensibly, a regulation that carries the force of law is legislative and therefore must satisfy notice and comment procedures. Whether a regulation receives *Chevron* deference also depends on whether it carries the force of law; but, confusingly, “it is not at all clear whether the force of law occupies precisely the same conceptual space” in these two different contexts. Kristin E. Hickman, *Unpacking the Force of Law*, 66 VAND. L. REV. 465, 467 (2013).

86. In an empirical study analyzing Treasury regulations promulgated between 2003 and 2005, Hickman found that among 137 tax regulations that cited general authority only (i.e., interpretative tax regulations), only 10 were promulgated without notice and comment, and at least half of the 10 claimed a good cause exception to notice and comment requirements. Kristin E. Hickman, *Coloring Outside the Lines: Examining Treasury’s (Lack Of) Compliance with Administrative Procedure Act Rulemaking Requirements*,

administrative law parlance.⁸⁷ Nonetheless, because my empirical methods rely on the pre-2011, tax exceptionalist⁸⁸ classification of some tax regulations as interpretative and some as legislative, this Article generally uses “interpretative” and “legislative” in that sense.

The interpretative/legislative distinction in tax law was not merely academic. For many years prior to *Mayo*, conventional wisdom also held that interpretative tax regulations would not receive full *Chevron* deference. Instead, they would receive less-deferential review under *National Muffler*, a case that pre-dated *Chevron* by five years. *National Muffler* addressed an interpretative regulation that the IRS promulgated using its general interpretive authority under section 7805(a) of the Code. The Supreme Court concluded that the regulation would warrant deference if it “implemented the congressional mandate in some reasonable manner.”⁸⁹ Reasonableness, in turn, was determined by a number of specific factors, including:

whether the regulation harmonizes with the plain language of the statute, its origin, and its purpose, [whether the regulation was a] substantially contemporaneous construction of the statute by those presumed to have been aware of congressional intent, . . . the manner in which it evolved, . . . the length of time the regulation has been in effect, the reliance placed on it, the consistency of the Commissioner’s interpretation, and the degree of scrutiny Congress has devoted to the regulation during subsequent re-enactments of the statute.⁹⁰

National Muffler’s factors emphasize statutory interpretation, making it directly comparable both to forms of judicial deference predating *Chevron* and potential replacements for *Chevron*. Judicial deference in the early twentieth century, as ultimately codified by the APA, simply awarded special weight to agencies’ contemporaneous construction of statutes and subsequent customary interpretive practices.⁹¹ These are the exact factors discussed in *National Muffler*. Moreover, many have suggested that the Court ought to return to a framework that

82 NOTRE DAME L. REV. 1727, 1751-53 (2007); see also *infra* notes 212-213 (discussing Hickman’s findings in more detail).

Although the IRS ultimately conducts notice and comment for almost all tax regulations, sometimes they are first promulgated as temporary regulations prior to notice and comment. *Infra* Appendix Section D.

87. See *infra* note 56 and accompanying text.

88. Cf. Kristin E. Hickman, *The Need for Mead: Rejecting Tax Exceptionalism in Judicial Deference*, 90 MINN. L. REV. 1537 (2006) (criticizing the tendency to treat tax regulations differently from other regulations). After *Mayo*, the IRS has tweaked its definitions of legislative and interpretative regulations to match other areas of law more closely, as reflected in the 2018 revision to its internal reference manual for employees. I.R.S., INTERNAL REVENUE MANUAL § 2.1.1.2.8 (Aug. 2, 2018).

89. Nat’l Muffler Dealers Ass’n v. United States, 440 U.S. 472, 476 (1979) (quoting United States v. Cartwright, 411 U.S. 546 (1973)).

90. *Nat’l Muffler*, 440 U.S. 472, 477 (1979).

91. Aditya Bamzai, *The Origins of Judicial Deference to Executive Interpretation*, 126 YALE L.J. 908, 916 (2017).

prioritizes statutory interpretation,⁹² making *National Muffler* a natural case study for the effect of replacing *Chevron*.

Because *National Muffler*'s factors are so specific, it was thought to be less deferential than *Chevron*, which set out a relatively vague standard for reasonableness. And because the facts in *National Muffler* concerned an interpretative regulation, most tax lawyers and judges believed that interpretative tax regulations continued to be subject merely to *National Muffler* deference, even after *Chevron*.⁹³ In contrast, legislative tax regulations were thought to receive full *Chevron* deference. The distinction mattered because interpretative tax regulations were, and remain, extremely common. A significant majority of tax regulations are interpretative in the sense in which tax lawyers used that term—during the period studied by this Article, 74.1% of all tax regulations cited section 7805 as their sole source of authority (meaning that they were interpretative), and only 0.014% of tax regulations cited some other section of the Code as their sole source of authority (meaning that they were legislative).⁹⁴ And regulations in general play an important role in tax law, especially since tax statutes are often written at a high level of generality that leaves substantial space for regulatory implementation.

In the decade before *Mayo*, the state of deference for interpretative tax regulations had become increasingly unclear.⁹⁵ Circuit courts had split on whether these regulations would receive full *Chevron* deference,⁹⁶ lesser *National Muffler* deference,⁹⁷ or some blurred combination of the two (sometimes on the basis that *Chevron* deference and *National Muffler* deference were indistinguishable⁹⁸). But the prevailing view was that interpretative tax regulations were subject to *National Muffler* deference rather than *Chevron*, and that *National Muffler* deference was indeed weaker.⁹⁹ As a result, the IRS would have expected that it

92. E.g., Murphy, *supra* note 76, at 3 (“[C]ourts reviewing agency statutory constructions should, contra *Chevron*, pick the constructions they deem best.”).

93. Hickman, *supra* note 88, at 1557 (“Although the practical difference is not always apparent, in [jurisdictions that accorded some Treasury regulations only *National Muffler* deference, rather than *Chevron* deference], specific authority regulations are given ‘controlling weight’ pursuant to *Chevron* while general authority regulations promulgated under I.R.C. § 7805(a) are given only ‘considerable weight’ under *National Muffler*.”). As Hickman has pointed out, this reasoning was hard to justify given that *Chevron* itself dealt with a regulation promulgated under the EPA’s general authority under the Clean Air Act, analogous to section 7805(a) of the Code. Hickman, *supra* note 86, at 1763-64.

94. *Infra* Appendix Section D.

95. Hickman, *supra* note 88, at 1556-59.

96. *Hosp. Corp. of Am. & Subsidiaries v. Comm’r*, 348 F.3d 136, 140-41 (6th Cir. 2003); *Swallows Holding, Ltd. v. Comm’r*, 515 F.3d 162 (3d Cir. 2008).

97. *Snowa v. Comm’r*, 123 F.3d 190, 197 (4th Cir. 1997); *Schuler Indus., Inc. v. United States*, 109 F.3d 753, 754-55 (Fed. Cir. 1997).

98. *Swallows Holding, Ltd. v. Comm’r*, 126 T.C. 96 (2006), *rev’d*, 515 F.3d 162; Hickman, *supra* note 88, at 1557-58.

99. E.g., Joana Que, *The State of Treasury Regulatory Authority After Mayo Foundation: Arguing for an Intentionalist Approach at Chevron Step One*, 85 S. CAL. L. REV. 1413, 1413-14 (2012) (noting that “[b]efore Mayo, Chevron did not have very much influence in the tax world,” and describing *National Muffler* as “a less deferential, tax-specific standard of review”).

would have to defend its interpretative regulations, at least in some courts, under *National Muffler* criteria.¹⁰⁰

The Supreme Court resolved this split in *Mayo*. *Mayo* concerned a regulation promulgated solely under the IRS's general section 7805 authority, and therefore was a clear example of an interpretative tax regulation.¹⁰¹ The taxpayer argued that the regulation should be given only *National Muffler* deference, rather than *Chevron* deference. But the Court disagreed. Justice Roberts, writing for a unanimous Court, failed to find "any justification for applying a less deferential standard of review to Treasury Department regulations than we apply to the rules of any other agency. In the absence of such justification, we are not inclined to carve out an approach to administrative review good for tax law only."¹⁰² He concluded that "*Chevron* and *Mead*, rather than *National Muffler* and *Rowan*, provide the appropriate framework" for evaluating interpretative tax regulations.¹⁰³

In other words, *Mayo* abolished the old, tax-specific distinction between interpretative and legislative regulations, at least for deference purposes. After *Mayo*, general administrative law standards applied to treat tax regulations as uniformly legislative, and therefore uniformly entitled to *Chevron* deference. By establishing that *Chevron* was the appropriate deference standard for interpretative tax regulations, the Supreme Court brought deference standards for these regulations back in line with those for regulations issued by other agencies. In doing so, it provided a unique opportunity for an empirical study on the impact of *Chevron* deference.

II. Data and Methods

A. Quantifying Text

Empirical scholars studying legal texts have traditionally read and coded individual documents: for example, by subjectively categorizing regulations based on their content.¹⁰⁴ But better datasets, faster computers, and modern techniques for text analysis have increasingly made it possible to use algorithmic natural language processing in place of human readers. Rather than holistic, subjective

100. Moreover, even prior to *Mayo*, some scholars had criticized the distinction between interpretative and legislative as fictive. Most prominently, Kristin Hickman argued that tax scholars and practitioners were indulging in "tax exceptionalism" when they attempted to carve out special categories of regulations solely within tax law. See Hickman, *supra* note 88. Instead, Hickman has argued (both before and after *Mayo*) that no Treasury regulations are truly interpretative within the meaning of the APA. See Hickman, *supra* note 86, at 1761-73. Hickman sparked a substantial scholarly re-evaluation of administrative tax law in the mid-2000s, by arguing that all tax regulations should receive *Chevron* deference. Hickman, *supra* note 88.

101. *Mayo Found. for Med. Educ. & Research v. Comm'r*, 562 U.S. 44, 56 (2011).

102. *Id.* at 55.

103. *Id.* at 57.

104. Frank Fagan, Book Review, *Natural Language Processing for Lawyers and Judges*, 118 MICH. L. REV. *13 (2021) (forthcoming) ("Older descriptive studies of legal doctrine that populate law reviews and treatises rely on hand-coded cases and small datasets.").

assessments, articles like this one quantify texts based on the specific words that they use.¹⁰⁵

For example, this Article uses terms like “fairness” and “efficiency” as a proxy for an agency’s normative orientation in promulgating rules. It then calculates the *term frequency* of these normative phrases by dividing the number of such phrases by the total word count of a particular text. A document with one hundred words and five normative terms would therefore have a normative term frequency of five percent. The Article similarly calculates the frequency of statutory terms by counting phrases like the “interpretation” of a “statute;” it calculates references to legislative history by counting references to legislative materials, like congressional reports and hearings; and it calculates references to public comments by counting references to “commenters.” Each set of proxies includes an extensive list of synonyms and ignores capitalization, punctuation, and stemming (word suffices like “-ly” or “-ing”). Section C of the Appendix describes the proxies in greater detail.

Some existing work on statutory interpretation by courts does not examine term frequency, but instead merely calculates the raw percentage of documents that have included particular terms—say, the percentage of appellate court cases each year citing legislative history—because this is all that has been feasible using Westlaw or Lexis searches.¹⁰⁶ Term frequency captures this nuance and more. The regression analysis in this Article asks not only of how many regulatory preambles cite certain concepts at all, but also how frequently preambles refer to those concepts when they do cite them.

In addition, using term frequency rather than a raw percentage of documents accounts for changes in the lengths of documents over time.¹⁰⁷ Hypothetically, if the IRS were to use one normative term every 100 words, then any increase in the length of preambles over time would increase the likelihood that any particular preamble contained a normative term. This would increase the percentage of preambles containing normative terms, but it would not increase term frequency.¹⁰⁸

105. A great deal of prior work has used either the count of relevant terms or their frequency as a methodological proxy. Choi, *supra* note 36; Aaron-Andrew P. Bruhl, *Statutory Interpretation and the Rest of the Iceberg: Divergences Between the Lower Federal Courts and the Supreme Court*, 68 DUKE L.J. 1 (2018); Corey Ditslear & James J. Brudney, *The Warp and Woof of Statutory Interpretation: Comparing Supreme Court Approaches in Tax Law and Workplace Law*, 58 DUKE L.J. 1231 (2009); Steven A. Dean & Lawrence M. Solan, *Tax Shelters and the Code: Navigating Between Text and Intent*, 26 VA. TAX REV. 879 (2007); John Calhoun, *Measuring the Fortress: Explaining Trends in Supreme Court and Circuit Court Dictionary Use*, 124 YALE L.J. 484 (2014).

106. Bruhl, *supra* note 105, at 30 (“[T]he analyses in this Article rely on electronic searches, primarily in Westlaw, to identify and count cases.”); Lawrence Solan, *Private Language, Public Laws: The Central Role of Legislative Intent in Statutory Interpretation*, 93 GEO. L.J. 453, 453-54 nn.118-19 (2005) (using Lexis searches to assess methodology).

107. Rachel Potter argues that agencies will tend to promulgate longer preambles when they expect pushback from the President, Congress, or courts. POTTER, *supra* note 45, at 97.

108. Note, however, that term frequency would not account for a nonlinear relationship between word count and term usage. For example, imagine that the first 1000 words of any given preamble do not include any

Term frequency also offers some advantages compared to more traditional hand-coding methods. It is less reliant on subjective reviewer evaluations and is therefore less prone to human error or intercoder unreliability.¹⁰⁹ And, because term frequency can be efficiently calculated by a computer, it allows studies to deal with large datasets of tens of thousands of texts (like the one used in this Article) that would be impracticable even for a dedicated team of human readers. Finally, it is more transparent and easier to replicate, because the judgments when calculating term frequency are primarily in the selection of terms rather than sophisticated and subjective evaluation of individual texts. All of the code used in this Article, in Python, Stata, and R, and the datasets generated for this Article, are publicly available online.¹¹⁰

Of course, term frequency has disadvantages as well. Most prominently, it is generally less nuanced in its evaluation of texts. “Fairness,” “efficiency,” and other terms used as proxies do not perfectly capture the qualitative nuances of any given agency’s explanation for its decision-making. As a result, term frequency can only capture broad changes that are perceptible across many different texts. Term frequency is also susceptible to changes in terminology—if (purely hypothetically) issues described in terms of “justice” fifty years ago were described in terms of “fairness” today, a simple analysis of the term “fairness” might suggest a change in attitude where none occurred. Finally, term frequency requires large bodies of high-quality data to be effective.

This Article attempts to mitigate these issues by focusing on high-level shifts in terminology, rather than more specific justifications for agency rulemaking. It also uses a relatively narrow 20-year snapshot in a time period for which high-quality official bulk data are publicly available, which lessens the risk of terminological change over time and quality issues in text analysis.

normative terms, but the next 1000 include many. If the IRS were to move from 1000-word preambles to 2000-word preambles, term frequency would still increase.

However, this hypothetical problem would be hard to square with the results in this Article, where normative terms and statutory terms move in opposite directions. That is, if normative terms increased solely as a mechanical function of word count increases, why did statutory terms decrease over the same period? Or, if statutory terms decreased solely as a mechanical function of word count decreases, why did normative terms increase over the same period?

109. Intercoder reliability refers to the likelihood that different coders will give the same score to the same document. It is especially important since in most studies only one coder will read each document, and several coders will typically be required to code a full dataset. *See generally* Matthew Lombard, Jennifer Snyder-Duch & Cheryl Campanella Bracken, *Content Analysis in Mass Communication: Assessment and Reporting of Intercoder Reliability*, 28 HUM. COMM. RES. 587 (2002) (discussing the concept of intercoder reliability).

110. *Code – How Did Chevron Affect Agency Rulemaking? An Empirical Study*, JONATHAN H. CHOI, jonathanhchoi.com/code-mayo-chevron (last updated Aug. 6, 2020) [<https://perma.cc/PE8W-4CNR>].

B. Preambles as Windows into the Regulatory Process

The APA requires any agency issuing a new regulation to also write a “statement of basis and purpose,” informally known as a preamble.¹¹¹ The preamble explains the agency’s rationale for the regulation and provides important guidance on how the regulation will apply. In addition, the preamble responds to any suggestions received in notice and comment. This Article studies the text of regulatory preambles in order to assess changes in rulemaking style. The obvious alternative would be to study the text of regulations themselves. However, regulatory text does not reveal whether an agency enacted a regulation for policy reasons or legal reasons, nor whether the regulation adopted, rebutted, or ignored public comments. Instead, regulatory text is primarily driven by the substantive subject matter that the agency wishes to address. Thus, preambles are a much more useful tool in considering shifts in agency attitudes toward rulemaking.

A certain kind of legal realist might argue that preambles do not accurately reflect agency priorities in drafting regulations. Agencies might, for example, determine policies totally independently from promises of judicial deference, and merely write preambles *ex post* using language selected to maximize the likelihood of receiving judicial deference. Under this theory, changes in deference regime would change the terms used in regulatory preambles, without changing the regulations themselves. If so, then preambles would be an inadequate substitute for regulations themselves, which are the real object of our interest.

This is a fair and fundamental critique of the methods in this Article. It parallels criticisms of empirical analysis of judicial opinions, which sometimes also posit that judges write opinions as *post hoc* rationalizations. On the other hand, regulatory preambles are an important part of the regulatory process—they must be adequately detailed in order for the regulation to be upheld¹¹² and are an important factor in judicial review of regulations.¹¹³ Moreover, they are frequently used as guides to the appropriate interpretation of regulations,¹¹⁴ so (like judicial opinions) they carry force even if they do not always tell a complete story.

First-person accounts from agency administrators agree that preambles reflect (at least partly) the regulatory process as it actually occurs.¹¹⁵ Ultimately, regulatory preambles are the best evidence that we have of the intentions of

111. The name derives from the requirement under the APA that any agency promulgating regulations issue “a concise statement of their basis and purpose.” 5 U.S.C. § 553(c) (2018).

112. *Indep. U.S. Tanker Owners Comm. v. Dole*, 809 F.2d 847, 853 (D.C. Cir. 1987); Stack, *supra* note 70, at 377 (“In general, the failure to issue an adequate statement of basis and purpose renders the agency’s action invalid.” (citing *Indep. U.S. Tanker Owners Comm. v. Lewis*, 690 F.2d 908, 920 (D.C. Cir. 1982))).

113. Stack, *supra* note 70, at 378-79.

114. *Id.* at 361 (“[T]he text of a regulation and its statement of basis and purpose [preamble] stand in a unique relationship: together, they constitute the act of regulation, an act that is not complete without either element of this couplet. Based on this premise, it does not make sense to interpret the text of a regulation independently from its statement of basis and purpose.”); Stack, *supra* note 70, at 1252 (describing preambles as “the most authoritative source of guidance about the meaning of agency regulations”).

115. Elliott, *supra* note 12, at 11.

regulators and their thought processes in promulgating regulations. Although this evidence is imperfect, it should be considered alongside other evidence in trying to form a complete picture of agency practice.

C. Difference in Differences

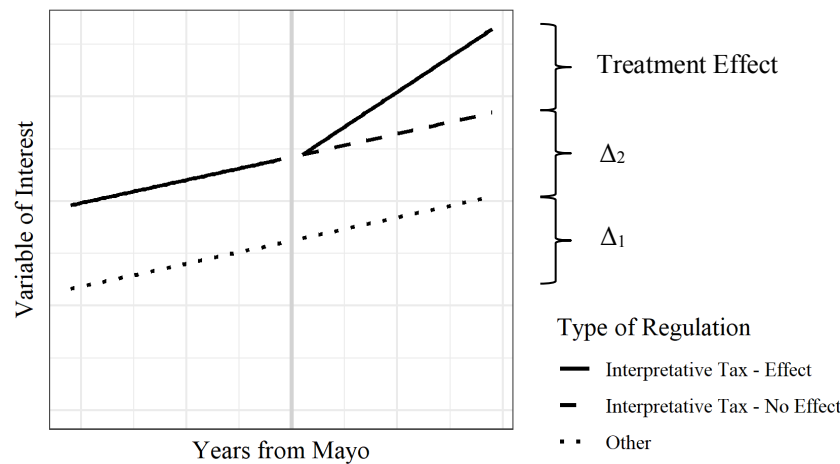
As noted above, the primary obstacle to any causal inference regarding the effect of *Chevron* is the difficulty of identifying treatment and control groups. *Chevron* theoretically applied across administrative law when it was decided in 1984.¹¹⁶ So, even if *Chevron* coincided with shifts in agency rulemaking styles, it would have been difficult to identify whether this was due to *Chevron* or to other contemporaneous events.

The primary contribution of this Article is to exploit a unique treatment and control group in order to test the effect of *Chevron* deference—interpretative tax regulations. Because *Mayo* was an unexpected event that altered the IRS’s rulemaking incentives, it can serve as the basis for a difference-in-differences study.¹¹⁷

A difference-in-differences model can be used to study the effect of any unexpected event that differentially affects a treatment and control group. Figure 3 below illustrates the basic difference-in-differences model, using interpretative tax regulations as the treatment group (solid line) and all other regulations as the control group (dotted line). The *y*-axis illustrates the variable of interest (say, frequency of the relevant term), and the *x*-axis specifies two periods, before and after *Mayo*. If *Chevron* deference has no effect, then there should be no difference between the dashed line and the solid line. A difference-in-differences model therefore tests whether the solid line is statistically significantly different from the dashed line.

116. *Chevron U.S.A. Inc. v. Nat. Res. Def. Council, Inc.*, 467 U.S. 837 (1984).

117. Phrased more technically, difference-in-differences analysis assumes that “in the pre-treatment the treatment had no effect on the pre-treatment population.” Michael Lechner, *The Estimation of Causal Effects by Difference-in-Difference Methods*, 4 FOUNDATIONS & TRENDS IN ECONOMETRICS 165, 178 (2010).

Figure 3: Hypothetical Difference-in-Differences Model

More technically, the model accounts for differences over time that apply both to the treatment and control groups (Δ_1 in Figure 3), as well as static differences between the treatment and control group (Δ_2 in Figure 3). It then attempts to identify any disproportionate effect of the unexpected event on the treatment group—the treatment effect, or “difference in differences.”

By focusing on the difference in differences, the model controls for pre-existing differences between the treatment and control groups, as well as static trends. Hypothetically, imagine that courts generally deferred more to tax regulations than other regulations as a matter of course, regardless of the applicable deference regime. It would consequently be misleading merely to compare the levels of normative or statutory term frequencies at any single point in time (either before or after *Mayo*), because courts might be more deferential to tax regulations during both periods. Similarly, imagine that pressure from different presidential administrations had gradually pushed all agencies to justify rulemaking in more normative terms. It would therefore be misleading solely to compare interpretative tax regulations before and after *Mayo*, because the shift in the style of interpretative tax regulations might be true of all regulations. By instead focusing on the differences in trends between the treatment and control groups, this method controls for level differences between the groups that persist over time, as well as time trends that apply to all regulations.

If critics of *Chevron* like Justices Thomas and Kavanaugh,¹¹⁸ as well as supporters of *Chevron* like Cass Sunstein,¹¹⁹ are correct, then the treatment effect should be positive for normative terms and negative for statutory terms, reflecting a shift toward normative and away from statutory explanations for rulemaking. Similarly, if my model predicting a link between heightened deference and

118. See *supra* note 4 and accompanying text.

119. See *supra* note 40 and accompanying text.

procedural effort is correct, then the treatment effect should be positive for proxies of procedural effort (the length of regulatory preambles and the frequency of references to commenters). If skeptics of *Chevron*'s influence are correct, then the treatment effect in both cases should be small or zero.

Difference in differences is a useful tool in causal inference, but it makes some significant assumptions as well, most prominently the assumption that the trends between the treatment and control groups would have remained parallel absent the treatment. It is impossible counterfactually to observe what would have occurred in the absence of the treatment (*Mayo*). However, Sections IV.F through IV.H apply several statistical techniques to assess the validity of the parallel trend assumption and explore alternative models in case the assumption is violated.

Beyond the basic example in Figure 3, additional control variables make the model more nuanced and accurate—for example, by adding time trends before and after *Mayo* in order to model changes year-over-year, rather than two blunt pre-*Mayo* and post-*Mayo* categories. The full model controls for linear time trends, separately estimated pre- and post-*Mayo*; it also controls for the type of regulatory document (rule, amendment, notice, or other), the length of the relevant regulation's text,¹²⁰ the reading level of the preamble,¹²¹ and the party of the president in power when the regulation was promulgated. Controlling for the type of regulatory document and the length of the associated regulation addresses bias from changes in the types of regulations issued over time, as well as complexity or length of underlying regulations; controlling for reading level addresses bias from changes in the overall complexity of preambles over time, which could potentially be correlated with interpretive methodology; and controlling for party of the current presidential administration addresses bias from shifts in regime and possible concomitant shifts in administrative priorities.

An additional nuance is that this Article uses a two-part regression model to analyze specific terms, meaning that it separately estimates the likelihood that a preamble will use *any* terms of a particular type (step one) and, *conditional* on using at least one term of that type, the quantity of terms used (step two).¹²²

The two-part model is used in order to more accurately model distributional features of the term frequency data, and the two steps are ultimately combined to produce estimates of average linear marginal effects, much like a conventional ordinary least squares (OLS) regression.¹²³ However, the two-step regression also has the benefit of more granularly modeling the effect of *Chevron*. For example,

120. Note that the regressions do not control for the length of the preamble, which is already used to calculate term frequency.

121. Reading level is measured using the Coleman-Liau readability index, which uses the average number of letters per word and the average number of words per sentence to approximate the grade level that would be appropriate for a given text. Meri Coleman & T.L. Liau, *A Computer Readability Formula Designed for Machine Scoring*, 60 J. APPLIED PSYCH. 283 (1975).

122. See *infra* Appendix Section G.

123. See *infra* Appendix Section G.

if a difference-in-differences coefficient were positive in the first step of the two-part regression, but zero in the second step, that would imply that *Chevron* makes an agency more likely to use any instances of that term, but does not affect the *intensity* of that term's usage.

D. Potential Confounders

A key assumption of the difference-in-differences model is that the treatment is the *only* event that differentially affects the treatment and control groups during the relevant period. This is a qualitative question of study design rather than a quantitative statistical question. To answer it, we need to consider other events in administrative tax law between 2000 and 2020, particularly those occurring around *Mayo* in 2011.

Mayo was a rare example of a Supreme Court tax case and marked a seismic shift in administrative tax law. Moreover, it was a surprise to many tax lawyers and scholars when it was decided,¹²⁴ making it a good candidate as a treatment in a difference-in-differences study. However, other court decisions and scholarly commentary over the past two decades could also have affected Treasury's regulatory drafting practices.

First, prior to *Mayo*, the lines between *Chevron* and *National Muffler* deference were not as crisp as one might like. The Supreme Court's 2001 decision in *Mead* provided an early indication that *Chevron* might be the appropriate deference standard for interpretative tax regulations.¹²⁵ Similarly, scholars have argued against tax exceptionalism both before and after *Mayo*.¹²⁶ This could have caused the IRS to begin anticipating heightened deference even prior to *Mayo*. As a result, the measured effect of *Mayo*, both in magnitude and statistical significance, might be underestimated by the model in this Article. That would suggest that the shift in deference regime could be even more consequential than this Article indicates.

Second, developments in APA jurisprudence after *Mayo* also might confound estimates of *Mayo*'s direct effects. These include recent Tax Court and Ninth Circuit rulings in *Altera Corporation v. Commissioner*, which dealt with *Chevron* and *State Farm* issues and which drew heavily on *Mayo*,¹²⁷ recent procedural changes, like OIRA review of Treasury regulations that began in 2018,¹²⁸ and broader developments in administrative law,¹²⁹ such as recent

124. See Patrick J. Smith, *Life After Mayo: Silver Linings*, TAX NOTES, June 20, 2011, at 1253.

125. See generally Hickman, *supra* note 88 (arguing, after *Mead* but prior to *Mayo*, that administrative tax law should follow rules generally applicable in other fields of administrative law).

126. For an early example, see Merrill & Watts, *supra* note 83, at 570-75. For a more recent example, see Kristin E. Hickman, *Administering the Tax System We Have*, 63 DUKE L.J. 1717 (2014).

127. 898 F.3d 1266 (9th Cir. 2018), *cert. denied*, 141 S. Ct. 131 (2020); 145 T.C. 3 (2015).

128. See Clinton G. Wallace, *Centralized Review of Tax Regulations*, 70 ALA. L. REV. 455, 478-81 (2018).

129. Moreover, the difference-in-differences model controls for developments in administrative law not specific to tax law.

Supreme Court cases that linked *Chevron* and *State Farm* more explicitly.¹³⁰ These developments would tend to cause *overestimation* of the causal effect of *Mayo*, which would be problematic for the model in this Article.¹³¹ Luckily, these confounders can be addressed by shortening the window of time subject to analysis to end prior to the confounding events. Section IV.D conducts this alternative analysis and finds estimates that remain statistically significant, and in some cases even greater in magnitude.

The third and most difficult category is any development approximately contemporaneous with *Mayo*. The most troubling candidate is *Cohen v. United States*,¹³² a D.C. Circuit ruling which held that IRS subregulatory guidance could constitute a “final agency action” reviewable by courts.¹³³ *Cohen* concerned subregulatory guidance promulgated without notice and comment, whereas *Mayo* concerned interpretative tax regulations subject to notice and comment.¹³⁴ Thus, *Cohen* should primarily affect IRS revenue rulings, notices, and other subregulatory documents that are not studied in this Article. Moreover, because it represented the decision of only one circuit on an unusually egregious set of facts,¹³⁵ its influence on broader IRS practice may have been limited. And, to the extent that *Cohen* was finally decided only after *Mayo* and cited *Mayo*,¹³⁶ it might more properly be considered an application of *Mayo* rather than a separate event.

But the broader point remains: court decisions, no matter how momentous or surprising, are not as exogenous as earthquakes and hurricanes. They are written in a context that includes a constant stream of new scholarship, court rulings, and administrative reform. This Article employs a variety of techniques to isolate the effect of the treatment on the treated, but its findings are only one piece of evidence and could be usefully supplemented by future studies.

E. Limitations

Part V employs various specification checks to buttress this Article’s central difference-in-differences model. But some challenges resist statistical checks. First, because this Article uses interpretative tax regulations as its treatment group, there is a question of external validity—how similar is the IRS to other

130. *Encino Motorcars, LLC v. Navarro*, 136 S. Ct. 2117, 2125 (2016); *Michigan v. EPA*, 135 S. Ct. 2699, 2706 (2015).

131. Note that these developments are only confounders if we treat them as independent events rather than consequences of *Mayo*. Given that they cite and are directly influenced by *Mayo*, it is not clear that this treatment is correct, but it is at least more cautious.

132. *In re Long-Distance Tel. Serv. Federal Excise Tax Refund Litigation*, 539 F. Supp. 2d 281 (D.D.C. 2008), *rev’d*, *Cohen v. United States*, 578 F.3d 1 (D.C. Cir. 2009), *aff’d en banc*, *Cohen v. United States*, 650 F.3d 717 (D.C. Cir. 2011).

133. *Cohen*, 578 F.3d at 12.

134. *Mayo Found. for Med. Educ. & Res. v. United States*, 562 US 44, 54 (2011).

135. *Cohen* concerned an excise tax that the IRS had erroneously collected and agreed to refund only for taxpayers who went through a “virtual obstacle course.” *Cohen*, 650 F.3d at 736. The D.C. Circuit noted that “[t]he litigation position of the IRS throughout the history of the excise tax has been startling.” *Id.*

136. *Cohen*, 650 F.3d at 736. The opposite is not true, that is, *Mayo* did not cite *Cohen*.

agencies? How closely can we expect the specific experience of the IRS to generalize if *Chevron* is entirely eliminated? Federal agencies are sprawling bureaucracies with deep institutional idiosyncrasies, perhaps so much so that no single agency can serve as a test case for others.

While no two agencies are exactly the same, it is reasonable to expect some similarities in how different agencies would respond to changes in the level of judicial deference. The causal stories presented in this Article for why judicial deference would change regulatory methods does not depend on quirks of the IRS, but rather on high-level incentives that one would encounter in any agency. While external validity is always a concern, events that affect only a subset of agencies remain our best means to evaluate the causal effect of deference regimes on regulatory activity.

Similarly, *National Muffler* was a standard of review specific to tax law—the more salient question for administrative law in general is not the difference between *National Muffler* and *Chevron*, but between *Skidmore* and *Chevron*. It is especially striking that *Chevron* would differ from *National Muffler*, since both require courts to defer to “reasonable” agency interpretations, whereas *Skidmore* deference is conceptually much further from *Chevron*. Thus, to the extent that *National Muffler* was an intermediate deference standard between *Skidmore* and *Chevron*, that would imply that the elimination of *Chevron* would be even more dramatic than the findings in this Article suggest.

Moreover, after *Mayo*, some doubt remains about whether Treasury regulations that become effective prior to notice and comment (like temporary regulations) should receive *Chevron* deference.¹³⁷ This could mean that post-*Mayo* regulations may not receive the degree of heightened deference that one might have anticipated, which would attenuate the estimated causal effect of *Mayo*. Or, inversely, the estimated treatment effect might be reduced if agencies other than the IRS were emboldened by *Mayo*. If these other agencies took *Mayo* as a signal of the Supreme Court’s interest in enforcing *Chevron*, then *Mayo* might have encouraged them to rely on *Chevron* more heavily than before. If so, then the control group would also have received some of the treatment effect, and the comparison between the treatment and control groups would not be as clean. Again, these problems would attenuate the measured effect of *Mayo* and therefore decrease the estimated effect of *Chevron*. That implies that the effect magnitudes and statistical significance of the results in this Article may be underestimated. Consequently, it may be safest to interpret these effect sizes as lower bounds in magnitude for the causal effect of *Chevron*.

137. Lederman, *supra* note 84, at 662-63; Hickman, *supra* note 86, at 1760 (“Treasury is not the only agency that promulgates binding regulations in advance of seeking and considering public comments. Nevertheless, the courts generally consider regulations issued through such a process procedurally invalid unless one of the four exceptions listed in APA section 553 applies. Many if not most Treasury regulations do not fall within the scope of those exceptions.”).

III. Results

A. *Statutory and Normative Terms*

To test the relative importance of legal and policy considerations in agency rulemaking, this Article examines the relative frequencies of normative terms—those referring to normative issues, like fairness and efficiency—and statutory terms—those referring to statutory interpretation.

Simple trends suggest that *Mayo* marked a significant shift away from statutory explanations for rulemaking and toward normative explanations. Figures 4 and 5 show the annual average term frequencies for normative and statutory terms,¹³⁸ along with trend lines¹³⁹ separating interpretative tax regulations from all other regulations. (The year-by-year figures in this Article generally exclude 2000 as a partial year.) These figures suggest an inflection point when *Mayo* was decided, after which interpretative tax regulations become more normative and less statutory relative to other regulations.

138. Averages are calculated as the number of terms of a particular type across all regulations issued that year, divided by the total word count of regulations issued that year. “Other” regulations include both non-interpretative tax regulations and non-tax regulations.

139. The smoothed trend lines throughout this Article are generated using locally estimated scatterplot smoothing, a non-parametric form of local regression that fits a smooth curve to data points. See WILLIAM S. CLEVELAND, *THE ELEMENTS OF GRAPHING DATA* 168-73 (rev. ed. 1994) (describing LOESS). All of the confidence intervals in this Article are calculated using bootstrapping. Given a sample of data points, bootstrapping recreates a sample of the same size by randomly sampling (with replacement) from the original sample. This is repeated a number of times, here 2000 times, and LOESS curves are recalculated with respect to each bootstrapped sample. For each point on the graph’s x-axis (here, each point in time), the values of each bootstrapped LOESS curve are stored and then used to calculate a confidence interval.

The confidence intervals follow the basic bootstrap (also known as the “reverse percentile,” “pivotal,” or “empirical” bootstrap) equation. The basic bootstrap calculates confidence intervals using some α such that the probability of an estimate lying within the confidence interval is $1-\alpha$. For example, given $\alpha = 0.05$ (a 95% confidence interval), then for any point on the x -axis, where θ is the LOESS value in the original sample, $\theta_{0.025}^*$ is the 2.5th-percentile bootstrapped value, and $\theta_{0.095}^*$ is the 97.5th-percentile bootstrapped value, the confidence interval equals:

$$(2\theta - \theta_{0.095}^*, 2\theta - \theta_{0.025}^*)$$

A. C. DAVISON & D. V. HINKLEY, *BOOTSTRAP METHODS AND THEIR APPLICATION* 194 (1997). Note that the confidence intervals are the confidence intervals of the *curve*, not confidence intervals of *observations*. That is, within each interval with respect to a given point on the x -axis, there is an estimated 95% probability that the true regression line lies within that interval. But this does *not* imply that there is a 95% probability that any observation will lie within that interval. The latter probability would be captured by a prediction interval, which would take into account both uncertainty regarding the regression line as well as pointwise variance in the distribution of observations.

See generally Choi, *supra* note 36, at 377 n.64 (providing an identical explanation of LOESS and bootstrapping in a similar context).

Figure 4: Frequency of Normative Terms in Regulatory Preambles, 2001-2020

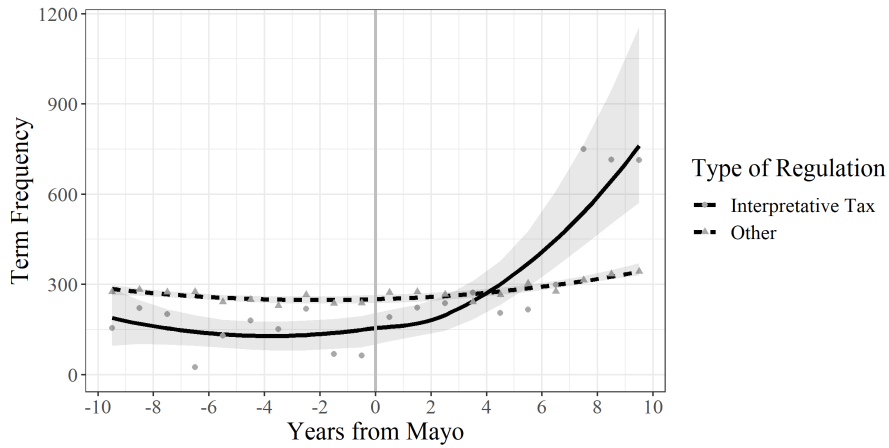
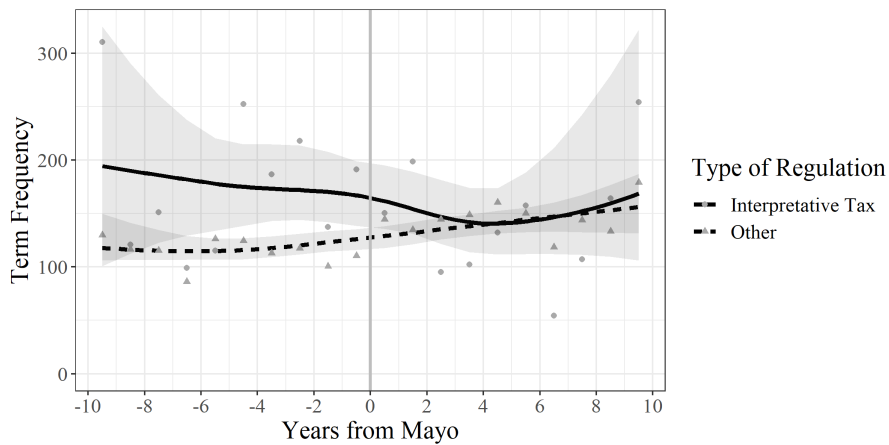


Figure 5: Frequency of Statutory Terms in Regulatory Preambles, 2001-2020



These graphs can be further supplemented with regression analysis. Regression analysis provides a specific estimate of the treatment effect and controls for potential confounding variables. The full list of control variables and more detailed methodologies are discussed in Section G of the Appendix.

Table 1 reports key findings from the regression analysis, which are laid out in more detail in Appendix Section G. Tables 14 and 15 in the Appendix contain full results from these regressions. All of the key results are statistically significant at a 99% confidence level. In raw numerical terms (rather than percentage terms), the average marginal effect of *Mayo* was to increase the frequency of normative terms by 154.4 per million words (against a baseline average of 186.6 normative terms per million across the entire sample), and to decrease the frequency of

statutory terms by 68.6 terms per million words (against a baseline average of 104.2 statutory terms across the entire sample).¹⁴⁰

Table 1: Estimated Treatment Effects for Normative and Statutory Terms, 2000-2020

Change in . . .	
Likelihood of Using Any Normative Terms	+126.8% ***
Term Frequency if Normative Terms Are Used	+22.6% ***
Overall Normative Term Frequency	+137.3% ***
Likelihood of Using Any Statutory Terms	-37.7% ***
Term Frequency if Statutory Terms Are Used	-20.9% ***
Overall Statutory Term Frequency	-48.5% ***

*** p<0.01, ** p<0.05, * p<0.1

Taken together, these findings suggest that the application of *Chevron* deference to interpretative tax regulations had a large and statistically significant effect on Treasury practice. These effects occurred along two different dimensions. First, relative to other regulations, *Mayo* triggered an additional increase in the likelihood that preambles of interpretative tax regulations would discuss normative issues at all. Second, among the preambles that use normative terms, *Mayo* triggered an additional increase in the frequency of normative terms that interpretative tax regulations use. This implies that the baseline likelihood of using normative terms comparatively increased alongside the *intensity* of those terms' use. Similarly, *Mayo* triggered a relative decrease in the likelihood that preambles of interpretative tax regulations would discuss statutory issues at all, and triggered a relative decrease in the frequency of statutory terms when they were used. Again, this implies a comparative decrease both in the baseline likelihood of any statutory discussion alongside a decrease in the intensity of such discussion.

The overall picture is remarkably consistent. *Mayo* marked a shift toward more normative and less statutory justifications for interpretative tax regulations, compared to other regulations. This played out both in the likelihood of using

140. Average marginal effects are calculated by estimating individual marginal effects per observation, and then taking the average of these estimated marginal effects across the entire sample. They should not be confused with marginal effects at the means, which are computed by subtracting the mean dependent variable values between two groups (e.g., interpretative tax preambles before and after *Mayo*).

those terms at all, and the intensity of usage when Treasury did use those terms. These changes were statistically significant and consistently large in magnitude.¹⁴¹

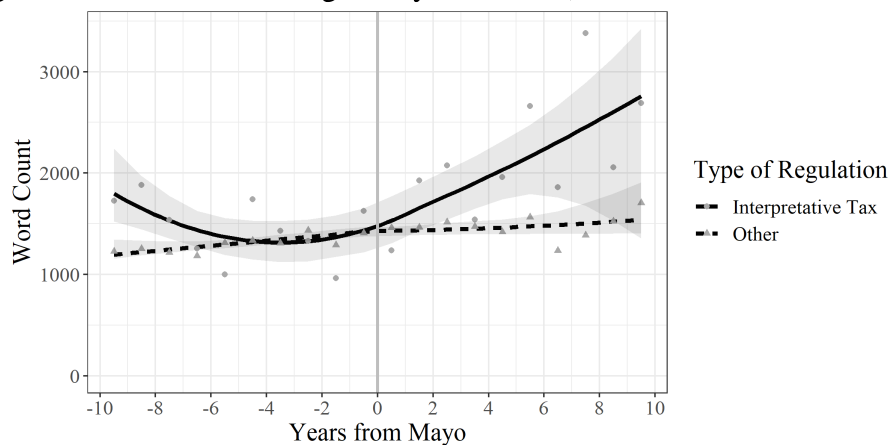
What does this mean for efforts to reform *Chevron*? In some respects, it underscores the claims of *Chevron*'s critics, who have long suggested that *Chevron* causes a shift in focus away from legal issues and toward policy issues. At the same time, many of *Chevron*'s supporters may be untroubled by these results, and may in fact find this effect of *Chevron* desirable. The findings most undercut those skeptics of *Chevron* who believe that it has little or no effect. Regardless of the ongoing debate over the application of *Chevron* by courts, this Article underscores that for agencies, *Chevron* matters a great deal.

B. Procedural Effort

This Article uses two different proxies of agency effort in complying with rulemaking procedures: preamble length and the frequency of references to commenters. *Ceteris paribus*, a longer preamble provides more detail and better satisfies the requirements of *State Farm* and the APA. (Because the regression analysis controls for regulation length, it specifically studies increases in preamble length holding regulation length constant.) Likewise, the frequency of references to commenters reflects an agency's relative interest in addressing public comments.

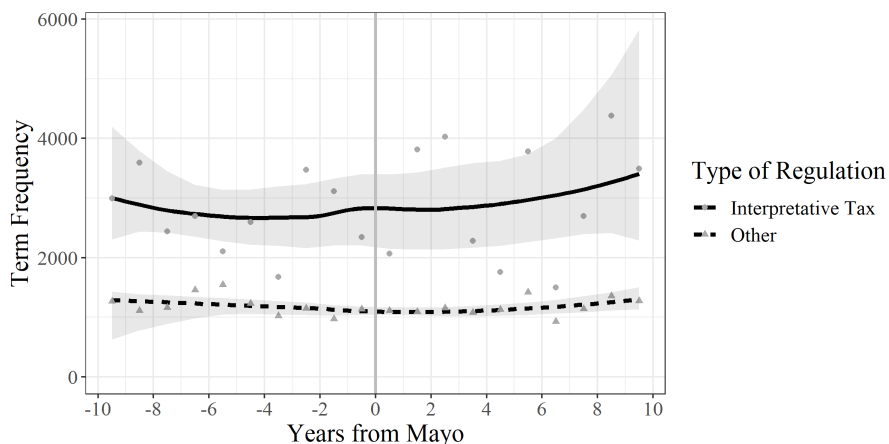
Figures 6 and 7 suggest that *Mayo* caused a disjunction in the length of interpretative tax preambles and the frequency of references to commenters. (The slope of the curve for preamble length prior to *Mayo* raises issues with the difference-in-differences model, which Sections IV.F through IV.H address in greater detail.)

Figure 6: Word Count of Regulatory Preambles, 2001-2020



141. Note that the changes are *relative* rather than absolute. The preambles to interpretative tax regulations will naturally still discuss statutes a great deal; this Article finds not that they discuss statutes little, but that they discuss them *less*.

Figure 7: Frequency of References to Commenters in Regulatory Preambles, 2001-2020



These trends are generally consistent with the hypothesis that heightened *Chevron* deference encourages agencies to exert additional rulemaking effort. However, note that the upward shift after *Mayo* in Figure 6 is much larger than in Figure 7. The magnitude of the shift in Figure 7 is very small and very noisy, suggesting that additional regression analysis is required.

Table 2 presents key regression results; the full results are in Tables 17 and 18 of the Appendix. Note that preamble length is examined in a single-step regression. This is because preamble length should never be zero, so the first-step regression is not required.

Table 2: Regression Results for Preamble Length and Frequency of References to Commenters, 2000-2020

Change in . . .	
Average Preamble Length	+18.9%**
Likelihood of Any References to Commenters	+1.1%
Term Frequency if Commenters are Referred to	+34.7%***
Overall Commenter Term Frequency	+35.8%*

*** p<0.01, ** p<0.05, * p<0.1

Regression analysis suggests that *Mayo* significantly increased the average length of preambles by 18.9%. The analysis also suggests that *Mayo* increased the frequency of references to commenters, albeit in a subtler way. Based on these results, *Mayo* did not significantly affect the likelihood of any particular preamble

referring to commenters; however, it significantly increased the intensity of discussion of comments by 34.7%. This latter finding makes intuitive sense. Some regulations only refer to public comments cursorily, or do so in passing while acknowledging the agency's obligations to consider comments under the APA.¹⁴² The term frequency of references to commenters conditional on at least one reference is a much better proxy for the depth of engagement with public comments. The overall estimated effect of *Mayo* on references to public comments is also large in magnitude, although because it is a noisier estimate, it is less statistically significant than the other results in this Article.

In non-percentage terms, the average marginal effect of *Mayo* was to increase preamble length by 250.3 words (against a baseline average of 1355.7 words across the entire sample) and to increase references to commenters by 183.5 words per million (against a baseline average of 599.8 words across the entire sample).

The link between *Chevron* deference and procedural effort finds additional support in the academic commentary that followed *Mayo*. Although *Mayo* only concerned the application of *Chevron* to an interpretative tax regulation and not Treasury's obligation to conduct notice and comment, some scholars drew the connection on their own: they argued that because *Chevron* deference only applied to regulations carrying the force of law, and because regulations carrying the force of law must be promulgated through notice and comment, any ruling that expanded the reach of *Chevron* among tax regulations necessarily expanded notice and comment obligations as well.¹⁴³

The Chief Counsel of the IRS noted shortly after *Mayo* that "great power bring[s] great responsibility," and that the new application of *Chevron* deference obligated the IRS to "make choices based on wise public policy," to "consider secondary and tertiary effects on stakeholders and the regulatory system writ large," and to "consult at length with affected internal and external parties."¹⁴⁴ Moreover, immediately after *Mayo*, the IRS revised its internal guidelines to drop the longstanding claim that "most IRS/Treasury regulations are interpretative, and therefore not subject to" procedural requirements under the APA.¹⁴⁵ Thus the scholarly commentary, statements from key IRS officials, and changes to IRS guidelines all support the view that procedural effort serves as the price of deference.

Procedural requirements play an important role in ongoing debates over the future of *Chevron*. Some scholars have suggested that *Chevron* should be modified to ask more deeply whether preambles provide adequate explanation

142. Updating OSHA Standards Based on National Consensus Standards; Signage, 78 Fed. Reg. 35,560 (June 13, 2013) (to be codified at 29 C.F.R. pts. 1910, 1926).

143. E.g., Richard W. Murphy, Pragmatic Administrative Law and Tax Exceptionalism, 64 DUKE L.J. ONLINE 21, 23 (2014); Michael Hall, From Muffler to Mayo: The Supreme Court's Decision to Apply Chevron to Treasury Regulations and Its Impact on Taxpayers, 65 TAX LAW. 695, 708-09 (2012).

144. IRS Chief Counsel Discusses Guidance, IRS Challenges, 2011 TAX NOTES TODAY 22-15 (Jan. 25, 2011).

145. Hickman, *supra* note 85, at 495-96 n.164.

for rulemaking. Catherine Sharkey, for example, recently proposed the incorporation of *State Farm* “reasoned decisionmaking” review into *Chevron*, requiring judges to take a “hard look” at the policy rationale for rulemaking rather than labelling any choice between two permissible interpretations of a statute per se reasonable.¹⁴⁶ Other administrative law scholars have suggested similar moves that substantively evaluate agency explanations for regulations as part of *Chevron* step two.¹⁴⁷

Most of these sources describe increased scrutiny during *Chevron* as normatively desirable, but not yet established doctrine. They also suggest that *Chevron* step two provides sufficient deference that agencies have little incentive to explain the policies underlying their rules.¹⁴⁸ But if agencies treat increased public engagement as the price of deference, it could be that one key benefit of more intense judicial review during *Chevron* step two—more careful deliberation by agencies during rulemaking—is already present under the status quo. That is, *Chevron* might already incentivize agencies to engage in a more thorough and responsive rulemaking process.

The evidence from this Article that *Chevron* increases procedural effort ultimately complements the recommendations of scholars attempting to bring hard-look review into *Chevron* step two. It suggests that wholesale elimination of *Chevron* would exacerbate existing complaints that agencies provide inadequate explanation for their rulemaking.¹⁴⁹ Thus, scholars hoping for a beefier rulemaking process should push for reform, rather than elimination, of *Chevron* deference.

C. Case Studies

Consider two case studies that concretely illustrate this Article’s quantitative findings. The first is a set of regulations promulgated in 2008, roughly three years prior to *Mayo*,¹⁵⁰ the second is a set of regulations promulgated in 2014, roughly three years after.¹⁵¹ These regulations illustrate a shift away from legal analysis and toward policy analysis, as well as an uptick in procedural effort. Both regulations cite section 7805 of the Code as their sole source of authority,¹⁵²

146. Sharkey, *supra* note 29.

147. *E.g.*, Seidenfeld, *supra* note 29; Levin, *supra* note 79.

148. *E.g.*, Seidenfeld, *supra* note 29, at 131 (“HHS, however, failed to explain how its interpretation responded to many of the policy concerns raised by the comments filed in the rulemaking proceeding.”).

149. *E.g.*, Sharkey, *supra* note 29, at 2365.

150. Suspension of Statutes of Limitations in Third-Party and John Doe Summons Disputes and Expansion of Taxpayers’ Rights To Receive Notice and Seek Judicial Review of Third-Party Summonses, 73 Fed. Reg. 23,342 (Apr. 30, 2008) [hereinafter *Summonses Regulations*].

151. Section 67 Limitations on Estates or Trusts, 79 Fed. Reg. 26,616 (May 9, 2014) [hereinafter *Section 67 Limitations*].

152. *Summonses Regulations*, *supra* note 150, at 23,342-44; *Section 67 Limitations*, *supra* note 151, at 26,616-18.

making them interpretative tax regulations. Both were final regulations issued following notice and comment on earlier proposed regulations.¹⁵³

The first, pre-*Mayo* set of regulations governs the summonses that the IRS issues to third parties in the course of tax assessments. The preamble to these regulations responded to extensive public comments. But the responses were relatively cursory, and Treasury rejected all of the suggestions received in the public comments. Moreover, it did so on exclusively statutory grounds, using familiar tools of statutory interpretation: reference to statutory text,¹⁵⁴ longstanding regulatory interpretations to which Congress was presumed to acquiesce,¹⁵⁵ the history of statutory amendments,¹⁵⁶ and the general structure of the statute.¹⁵⁷

The second, post-*Mayo* set of regulations governs the deductibility of costs incurred by trusts and estates.¹⁵⁸ This preamble also responded to public comments, but in much greater detail—although the underlying regulations were roughly a quarter the length of the earlier set, the preamble was slightly longer.¹⁵⁹ This time around, Treasury was both more responsive to public comments and more policy-focused. It accepted many of the suggestions proposed by commenters.¹⁶⁰ Moreover, the more recent preamble emphasized normative justifications for its rulemaking. In response to a comment arguing that Treasury lacked statutory authority to promulgate the regulations, the preamble described various grounds on which the regulation was normatively desirable: that it “provides equitable tax treatment to similarly situated taxpayers” and “reduces administrative burdens.”¹⁶¹ These rationales reflect a shift toward policy and away from law, and they are consistent with the high frequency of normative terms in

153. *Summonses Regulations*, *supra* note 150, at 23,343; *Section 67 Limitations*, *supra* note 151, at 26,616.

154. *Summonses Regulations*, *supra* note 150, at 23,343 (emphasizing that the regulations were “consistent with the language” of the statute); *id.* at 23,344 (arguing that a commenter’s suggestion “goes beyond the statutory language and the IRS’s authority to promulgate regulations”).

155. *Id.* at 23,343 (“Nothing in the statutory amendments . . . since these regulations were promulgated suggests that Congress intended to change” the meaning of “longstanding regulations” already issued.); *see also* Anita S. Krishnakumar, *Longstanding Agency Interpretations*, 83 *FORDHAM L. REV.* 1823, 1879 (2015) (discussing “the prevailing wisdom that longstanding agency statutory interpretations should receive heightened judicial deference”).

156. *Summonses Regulations*, *supra* note 150, at 23,343 (describing language changes in a 1998 statutory amendment); *id.* at 23,344 (“[T]hese regulations are interpretative of statutory provisions that have existed as law for several years”).

157. *Id.* at 23,344 (dismissing a suggestion as unnecessary in light of “the statutory structure”).

158. *Section 67 Limitations*, *supra* note 151, at 26,616.

159. The first set of regulations was 5325 words long, with a preamble of 2067 words; the second set was 1308 words long, with a preamble of 2439 words. It is possible that any increase in procedural effort after *Mayo* occurred at the cost of actually writing regulations themselves; this would be an interesting question for future research.

160. *Id.* at 26,616-18.

161. *Id.* at 26,618. It also cited a relevant case, but did not engage in conventional statutory interpretation. *Id.*

the later regulatory preamble, in contrast with the high frequency of statutory terms in the earlier regulatory preamble.¹⁶²

IV. Specification Checks

This Part discusses a number of tests that validate the model used in this Article and check its sensitivity to variations in specifications.

A. Reading Preambles to Confirm Proxy Terms

To confirm the validity of the terms chosen as proxies for discussion of normative issues, statutory issues, public comments, and legislative history, I randomly selected 160 regulatory preambles containing instances of these terms (40 preambles for each category, 20 from the period 2000–2009 and 20 from the period 2010–2019), out of the dataset used in this Article.¹⁶³ I then manually reviewed each preamble to check that the terms were used as expected. Upon review, I found eight instances where the term did not accurately refer to the concept it was meant as a proxy for, whereas in the other 595 instances, the terms were used as intended.¹⁶⁴ The overall false positive rate in the sample was therefore 1.3%.¹⁶⁵

The full list of preambles, including the specific terms used and full citations to the regulations reviewed, along with explanations where terms were not used as expected, is available online.¹⁶⁶

B. Regulatory Selection Effects

Another potential issue is the possibility of selection effects in the agency's choice between regulations and subregulatory guidance. We could imagine a simple model in which an agency seeking to promulgate any given policy chooses whether to do so in a regulation, subject to *Chevron* deference or *National Muffler* deference, or in subregulatory guidance, subject to *Skidmore* deference. Any change in the level of deference accorded to regulations might affect not just the agency's approach to the regulations it otherwise would have promulgated,

162. These regulatory preambles were selected as representative of the relevant shifts—the earlier regulatory preamble has 5 statutory terms and 0 normative terms out of 2067 words, whereas the later regulatory preamble has 4 normative terms and 0 statutory terms out of 2439 words.

163. The preambles included both tax and non-tax preambles; because non-tax preambles substantially outnumber tax preambles, the majority of the randomly selected preambles were non-tax.

164. There were 603 instances of terms in total: 50 normative terms from the 2000s, 77 normative terms from the 2010s, 34 statutory terms from the 2000s, 39 statutory terms from the 2010s, 89 commenter terms from the 2000s, 212 commenter terms from the 2010s, 75 legislative history terms from the 2000s, and 27 legislative history terms from the 2010s.

165. For consistency, the false positives manually identified were not dropped from the sample.

166. *Online Appendix: Randomly Selected Preambles to Confirm Term Frequency Results*, JONATHAN H. CHOI, jonathanhchoi.com/s/Randomly-Selected-Preambles.pdf (last updated Aug. 6, 2020) [<https://perma.cc/P6HD-7FTT>].

but might also cause the agency to substitute away from subregulatory guidance and toward regulations. Thus, even if an agency's approach to any particular policy remains the same, the average content of regulations might shift solely as a result of this substitution.

Do selection effects drive the results in this Article? One reason to suspect not is that the IRS has long been constrained in the resources that it can assign to draft regulations. All published IRS guidance, both regulatory and subregulatory, is issued by the Office of Associate Chief Counsel.¹⁶⁷ The IRS publishes a Priority Guidance Plan each year describing which regulations it will attempt to promulgate,¹⁶⁸ but many issues flagged by taxpayers will not be addressed in regulations for years. Moreover, because guidance is frequently time-sensitive, Treasury often issues subregulatory guidance in advance of regulations, either to clarify the law for taxpayers¹⁶⁹ or to head off potentially abusive transactions.¹⁷⁰ All of these constraints will limit Treasury's ability to freely select between types of guidance.

In addition, if substitution were to occur toward an increased number of formal regulations, we would expect Treasury to issue a greater proportion of regulations to subregulatory guidance after *Mayo*. Figure 8 graphs the total number of tax regulations¹⁷¹ divided by the total number of revenue rulings promulgated each year, from 2000 to 2020. It shows the opposite trend—that the ratio actually declined after *Mayo*.¹⁷²

167. I.R.S., INTERNAL REVENUE MANUAL § 32.1.1.1(4) (Nov. 13, 2019) (“Associate Chief Counsel offices are solely responsible for issuing published guidance.”). Note, however, that “on some projects, members of Operating Divisions may be involved in the development of a project.” *Id.* § 32.1.1.4.4(1) (Aug. 2, 2018).

168. I.R.S., INTERNAL REVENUE MANUAL § 32.1.1.4.1 (Aug. 2, 2018).

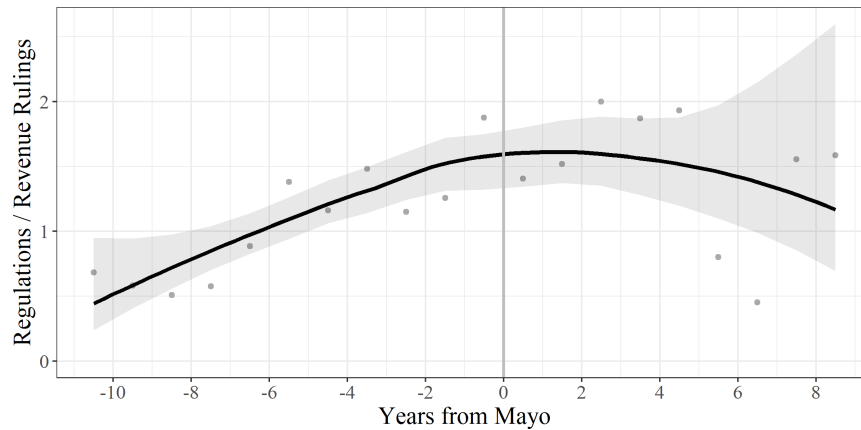
169. For example, Notice 2018-76 provided guidance in advance of Proposed Regulation 100814-19 on the deductibility of business meals. I.R.S. Notice 2018-76, 2018-42 I.R.B. 599; Meals and Entertainment Expenses Under Section 274, 85 Fed. Reg. 11,020 (proposed Feb. 26, 2020) (to be codified at 26 C.F.R. §§ 1.274-11, 1.274-12).

170. For example, Notices 2014-52 and 2015-79 provided guidance in advance of Proposed Regulation 135734-14 on the treatment of inversion transactions. I.R.S. Notice 2014-52, 2014-42 I.R.B. 712; I.R.S. Notice 2015-79, 2015-49 I.R.B. 775; Rules Regarding Inversions and Related Transactions; Notice of Proposed Rulemaking by Cross-Reference to Temporary Regulations, 82 Fed. Reg. 5476 (Jan. 18, 2017) (to be codified at 26 C.F.R. §§ 1.7874-7, 1.7874-10).

171. This includes both interpretative and legislative regulations.

172. The absolute number of tax regulations issued per year also declined after *Mayo*.

Figure 8: Total Tax Regulations / Revenue Rulings Per Year, 2000-2020



If anything, the decline in the ratio of regulatory-to-subregulatory guidance might support the ossification hypothesis—that, in order to spend greater time and effort promulgating each regulation, Treasury counterintuitively decreased the total number of new regulations that it issued each year. On the other hand, this decline in new regulations may also have been attributable to Trump-era executive orders generally making it more difficult for agencies to promulgate new regulations,¹⁷³ or budget cuts forcing the IRS to favor subregulatory guidance (which is issued without resource-intensive notice and comment). While the precise explanation for the substitution toward subregulatory guidance exceeds the scope of this Article, it would be an interesting topic for future research.

C. Legislative History as an Alternative Proxy for Statutory Terms

The frequency of statutory terms, as used in this Article, is calculated based on the number of times that an agency discusses statutes at all, for example by discussing how a “statute” should be “construed.”¹⁷⁴ An alternative and more specific proxy for statutory discussion is discussion of legislative history—for example, the frequency of citations to Senate reports, House floor debates, and other forms of legislative history. The term frequency of legislative history can be used as a dependent variable in a regression in the same way as the term frequency of normative or statutory terms.

173. See, e.g., Exec. Order No. 13,771, 82 Fed. Reg. 9339 (Jan. 30, 2017) (directing agencies to repeal two existing regulations for every new regulation).

174. See *infra* Appendix Section C.2. For purposes of defining statutory terms, this Article does not distinguish between interpretation and construction. See generally Lawrence B. Solum, *The Interpretation-Construction Distinction*, 27 CONST. COMMENT. 95 (2010) (discussing the distinction).

Because legislative history is the tool of statutory interpretation most commonly used by agencies,¹⁷⁵ we would expect it to follow the same trend as statutory terms in general. That is, we would expect the treatment effect to be negative with respect to legislative history, as with statutory terms. Table 3 shows the results of a regression studying the term frequency of legislative history, excerpted from Table 19 in the Appendix.¹⁷⁶ As expected, the overall treatment effect is negative, and the decline in the likelihood of any citations to legislative history is statistically significant at the 99% level. However, the intensity of discussion of legislative history essentially does not change. This suggests that the more reliable effect of heightened deference may be to decrease the likelihood of any statutory discussion at all.

Table 3: Regression Results for Legislative History Terms, 2000-2020

Change in . . .	
Likelihood of Using Any Legislative History Terms	-23.0%***
Term Frequency if Legislative History Terms Are Used	+2.4%
Overall Legislative History Term Frequency	-20.9%*

*** p<0.01, ** p<0.05, * p<0.1

In theory, another proxy for statutory references might be textualist terms—for example, citations to dictionaries and certain canons of interpretation, like language canons, which are favored by textualists.¹⁷⁷ However, as I noted in a previous article, the IRS has remained adamantly purposivist and uses virtually no distinctively textualist terms (like language canons or dictionaries) in its preambles, despite the rise of textualism at the Tax Court and other federal courts.¹⁷⁸ The dataset used in this Article reveals that this finding generalizes to other agencies as well: textualist terms are so rare that they cannot be statistically analyzed using term frequency.

D. Shortened-Window Regression Analysis, 2006-2015

Figures 4 and 5 suggest a general trend that the regression analysis supports, that *Mayo* caused a break in Treasury's approach to promulgating interpretative tax regulations. However, visual inspection suggests that the results may be

175. Choi, *supra* note 36, at 397-98 (describing how the IRS generally relies on legislative history to the exclusion of textualist interpretive tools).

176. Table 19 in the Appendix provides additional details from this regression.

177. Choi, *supra* note 36, at 384 n.90.

178. *Id.* at 397-98. Of course, the IRS frequently pays great attention to statutory text, but this is not a distinctly textualist technique, since purposivists also prioritize statutory text as important evidence of statutory meaning.

disproportionately driven by changes within the last few years, especially the increase in normative terms since 2018 shown in Figure 4. This is concerning, because it might cause the treatment effect to be conflated with other important recent events in tax administration—for example, the focus of the Trump administration on issues of fairness and efficiency,¹⁷⁹ or 2017 tax reform,¹⁸⁰ which may have caused a qualitative change in the type of regulations that Treasury issued in the most recent years (because they are issued with respect to new legislation).

To address this concern, I replicate the regression analysis in Part C for a shorter window of time that only includes the five years before and after *Mayo*: 2006–2015. Shrinking the window of time tests the robustness of the model to changes in specification and avoids outliers that may have been driven by recent events.

Table 4: Regression Results for Normative Terms, Statutory Terms, Preamble Length, and References to Commenters, 2006-2015

Change in . . .	
Likelihood of Using Any Normative Terms	+55.6% **
Term Frequency if Normative Terms Are Used	+14.3% **
Overall Normative Term Frequency	+63.7% **
Likelihood of Using Any Statutory Terms	-61.2% ***
Term Frequency if Statutory Terms Are Used	-21.5% ***
Overall Statutory Term Frequency	-66.5% ***
Average Preamble Length	+12.0% ***
Likelihood of Any References to Commenters	-3.4%
Term Frequency if Commenters are Referred to	+20.7% ***
Overall Commenter Term Frequency	+17.2%

*** p<0.01, ** p<0.05, * p<0.1

Full details from this regression are available in Table 20 in the Appendix. The overall results of the regression with a reduced window of time are similar. Table 4 shows that the sign of the coefficients for the key variable of interest are largely the same, and the results remain statistically significant at a 95% level or higher. The magnitudes of most of the findings are slightly smaller, except for the findings regarding statutory terms, which are larger than the baseline values in

179. Exec. Order No. 13,789, 82 Fed. Reg. 19,317 (Apr. 21, 2017) (directing agencies to promote regulations that are “simple, fair, efficient, and pro-growth”); *see also* Exec. Order No. 13,777, 82 Fed. Reg. 12,285 (Feb. 24, 2017) (requiring agencies to undertake reforms intended to “lower regulatory burdens on the American people”).

180. An Act to Provide for Reconciliation Pursuant to Titles II and V of the Concurrent Resolution on the Budget for Fiscal Year 2018, Pub. L. No. 115-97, 131 Stat. 2054 (2017).

Table 1. Because agencies take time to respond to changed incentives after court decisions, one would expect the treatment effect to be greater the more time has passed, and the decrease in magnitude is in line with this expectation. Moreover, because regulations can take years from proposal to completion, we should expect some lag before the effects of *Chevron* are fully visible.

The shortened-window findings are especially reassuring because of recent developments in administrative tax law that otherwise might have biased the results in this Article: *Altera*, *Encino Motorcars*, *Michigan v. EPA*, and reforms to OIRA review of tax regulations, as discussed above.¹⁸¹ This specification check suggests that the findings in this Article are not merely due to confounders.

E. Winsorized Regression Analysis

An alternative method that reduces the influence of outliers is to “winsorize” the dataset prior to conducting the regression, by replacing extreme values past some cutoff with values closer to the median. For purposes of this Article, I winsorize by taking the bottom 2.5% of values and replacing them with the 2.5th-percentile value, and likewise by taking the top 2.5% of values and replacing them with the 97.5th-percentile value.¹⁸² The net effect is to reduce the magnitude of the outliers and therefore to reduce their prominence in the regression results.

Table 5 excerpts the results of the winsorized regressions, while Table 21 in the Appendix contains full winsorized results. The key findings in this Article remain statistically significant. In addition, the magnitudes of the coefficients generally are smaller in the winsorized regressions; this is a typical consequence of winsorizing right-tailed datasets, which, while useful as a robustness check, does not produce point estimates that are preferable to the original specifications.

181. See *supra* notes 127-130 and accompanying text.

182. The most common parameters for winsorizing are 95% (the one used in this Article) and 90%. Because so many of the term frequency values are zero, *infra* Appendix Section F, winsorizing the left tail has no effect, but winsorizing the right tail has a substantial effect on point estimates.

Table 5: Winsorized Regression Results, 2000-2020

Change in . . .	
Likelihood of Using Any Normative Terms	+126.8%***
Term Frequency if Normative Terms Are Used	+13.3%***
Overall Normative Term Frequency	+119.2%***
Likelihood of Using Any Statutory Terms	-37.7%***
Term Frequency if Statutory Terms Are Used	-13.4%***
Overall Statutory Term Frequency	-43.6%***
Average Preamble Length	+15.0%**
Likelihood of Any References to Commenters	+1.1%
Term Frequency if Commenters are Referred to	+23.6%***
Overall Commenter Term Frequency	+24.7%

*** p<0.01, ** p<0.05, * p<0.1

F. Placebo Tests

As noted above, a key assumption of the difference-in-differences model is that the treatment and control groups would have followed parallel trends absent the treatment. Because we cannot observe this counterfactual, the standard alternative is to ensure that the treatment and control groups at least followed parallel trends *before* the treatment. We can do this by visually inspecting pre-treatment trends (pre-trends) to confirm that they were parallel prior to 2011.

Figures 4 and 5 show that the pre-trends were similar between the treatment and control groups for normative and statutory terms. Figure 7 is more equivocal, but does not reveal obvious differences in pre-trends between the treatment and control groups. However, Figure 6 suggests that the pre-trends substantially differed for preamble length; the treatment group pre-trend sloped downward, while the control group pre-trend sloped upward, making the reversal after *Mayo* even sharper.

Another test of the parallel trend assumption is the placebo test. Conventionally, a placebo test is conducted by excluding the post-treatment period, cutting the pre-treatment period in half, and assigning the second half of the pre-treatment period as a “placebo” treatment period. To conduct the placebo test, I exclude observations after *Mayo* and analyze the period 2000–2010,¹⁸³ assigning 2000–2005 as the pre-treatment period and 2006–2010 as the post-treatment period. The results of the placebo test are in Table 6 below (excerpted from Table 22 in the Appendix), from regressions with full controls:

183. The pre-treatment period used for the placebo test ends six months before *Mayo* is decided, to exclude anticipatory effects from briefs and oral arguments in the *Mayo* case that preceded the decision itself.

Table 6: Placebo Regression Results for Normative Terms, Statutory Terms, and Preamble Length, 2000-2020

Change in . . .	
Likelihood of Using Any Normative Terms	+14.7%
Term Frequency if Normative Terms Are Used	-0.5%
Overall Normative Term Frequency	+11.3%
Likelihood of Using Any Statutory Terms	+26.5%
Term Frequency if Statutory Terms Are Used	+2.7%
Overall Statutory Term Frequency	+27.6%
Average Preamble Length	-15.0%***
Likelihood of Any References to Commenters	+11.6%
Term Frequency if Commenters are Referred to	+14.8%
Overall Commenter Term Frequency	+25.5%

*** p<0.01, ** p<0.05, * p<0.1

The only statistically significant result from the placebo test is a decline in preamble length. This is consistent with the violation of the parallel trend assumption visible in Figure 6. The other results are both statistically insignificant and relatively small in magnitude, compared to the results in the baseline model. The results of the placebo test cannot confirm that the parallel trend assumption is satisfied—the insignificance of the results could be explained by low statistical power as well as by the satisfaction of the parallel trend assumption—but they provide some additional reassurance in that direction. The exception, of course, is the result of the placebo test with respect to preamble length, which must be addressed in another way.

G. Extrapolating Pre-Trends

If the parallel trend assumption is violated, difference in differences will produce biased estimates and cannot be relied upon. The estimates might still be directionally correct if the pre-trends could merely be extrapolated into the post-treatment period. If so, then the effect of *Mayo* on preamble length would merely be underestimated under conventional difference-in-differences analysis, because *Mayo* not only set the length of interpretative tax preambles on an upward trend but also reversed those preambles' prior downward trend. Moreover, the difference-in-difference estimates could be corrected by extrapolating the differences in pre-trends to the post-treatment periods, a common approach in existing literature.¹⁸⁴

184. See, e.g., Carlos Dobkin et al., *The Economic Consequences of Hospital Admissions*, 108 AM. ECON. REV. 308 (2018); Manudeep et al., *Broadband Internet: An Information Superhighway to Sex Crime?*, 80 REV. ECON. STUD. 1237, 1257 (2013). Following Manudeep et al., I extrapolate pre-trends by measuring the

Extrapolating pre-treatment time trends in the preamble length model predictably increases the magnitude of the estimated treatment effect. It increases the point estimate for the percentage change in average preamble length from 18.9% to 29.6%. The full results from this regression are presented in Table 17 in the Appendix.

Table 7: Regression Results for Preamble Length, Extrapolating Pre-Trends, 2000-2020

Change in . . .	
Average Preamble Length	+29.6%***
*** p<0.01, ** p<0.05, * p<0.1	

However, the assumption that pre-treatment time trends will persist into the post-treatment period is a strong one. A skeptic might instead suggest that the explanation was mean reversion—that some omitted variable caused a large but temporary drop in the length of IRS preambles prior to *Mayo*, which merely disappeared as IRS preamble lengths reverted to the mean after *Mayo*. If so, the effect of this omitted variable would be falsely attributed in the baseline model (even more so in the model extrapolating pre-trends) to *Mayo*. This Article cannot fully address this possibility; it provides reason to remain skeptical of the results regarding preamble length, and suggests that the price-of-deference model would benefit from further research.

H. Synthetic Controls

Another method to address violation of the parallel trend assumption is to use an entirely different model: synthetic controls. The synthetic control method does not assume that the treatment and control groups simply would have followed parallel trends absent the treatment. Instead, it uses a weighted combination of other agencies to generate a synthetic control group whose behavior is intended to approximate the behavior of the actual treatment group.

slope of preamble length relative to year for the treatment and control groups for years before 2011. Given the slope estimate v_g , for group g , where g is either the treatment group (interpretative tax regulations) or the control group (other regulations), I specify the following model:

$$\begin{aligned} \log[\text{Preamble Length}_i] &= \beta_0 + \beta_1 \cdot \text{PostMayo}_i + \beta_2 \cdot \text{Interpretative Tax Reg}_i + \beta_3 \cdot \text{PostMayo}_i \\ &\quad \cdot \text{Interpretative Tax Reg}_i + \beta_4 \cdot \text{Year}_i + \beta_5 \cdot \text{Year}_i \cdot \text{PostMayo}_i + \beta_6 \\ &\quad \cdot \text{Regulation Length}_i + \beta_7 \cdot \text{High Reading Level}_i + \beta_8 \\ &\quad \cdot \text{Democratic Administration}_i + \beta_9 \cdot \text{Rule}_i + \beta_{10} \cdot \text{Notice}_i + \beta_{11} \cdot \text{Year}_i \cdot v_g + \epsilon_i \end{aligned}$$

Cf. Manudeep, *supra*, at 1257 (using an analogous model).

The synthetic control group is generated by taking a weighted average of the actual control groups (called the “donor groups”), as well as covariates, to minimize some loss function (usually mean squared prediction error¹⁸⁵) with respect to the difference between the synthetic control and the treatment group for the pre-treatment period. The synthetic control group is then compared with the treatment group for the post-treatment periods, and any difference in the relevant dependent variable between the two groups is inferred to be attributable to the treatment.¹⁸⁶

Synthetic controls diverge from difference in differences in a variety of ways, two of which are particularly relevant here. First, synthetic control analysis is conducted on panel data where each observation is average term frequency or preamble length for a particular agency for a particular year. In contrast, the difference-in-differences analysis in this Article uses individual regulations as observations, meaning that the difference-in-differences regressions consider a much larger N than the synthetic control analysis. Second, as has been extensively discussed above, difference in differences assumes parallel trends. While the synthetic control method involves its own set of assumptions,¹⁸⁷ it crucially does *not* assume parallel trends.

Figures 9 through 12 below compare estimated term frequencies and preamble lengths for the actual IRS and a synthetic IRS.¹⁸⁸ In most years prior to *Mayo*, the point estimates for the IRS and the synthetic IRS are relatively close together, diverging in the expected directions after *Mayo*.¹⁸⁹

185. See Alberto Abadie, *Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects*, J. ECON. LIT. at *8-*9 (forthcoming 2021).

186. All of the synthetic control analysis in this Article was conducted in R, using the Gsynth package published by Yiqing Xu and Licheng Liu. Yiqing Xu & Licheng Liu, *Gsynth: Generalized Synthetic Control Method*, https://yiqingxu.org/software/gsynth/gsynth_examples.html (last updated Mar. 6, 2020) [<https://perma.cc/HM6H-3SS9>]. Gsynth was used because of its ability to process unbalanced datasets.

187. Abadie, *supra* note 185, at *25-*32.

188. Strictly speaking, the treatment group is the group of interpretative tax regulations, since the IRS also issues legislative regulations that are not part of the treatment group for purposes of this analysis. These group names are adopted for greater simplicity.

189. Again, the trend lines are produced through locally estimated scatterplot smoothing. The confidence intervals are calculated using $\alpha = 0.1$.

Figure 9: IRS v. Synthetic IRS, Normative Term Frequency, 2000-2020

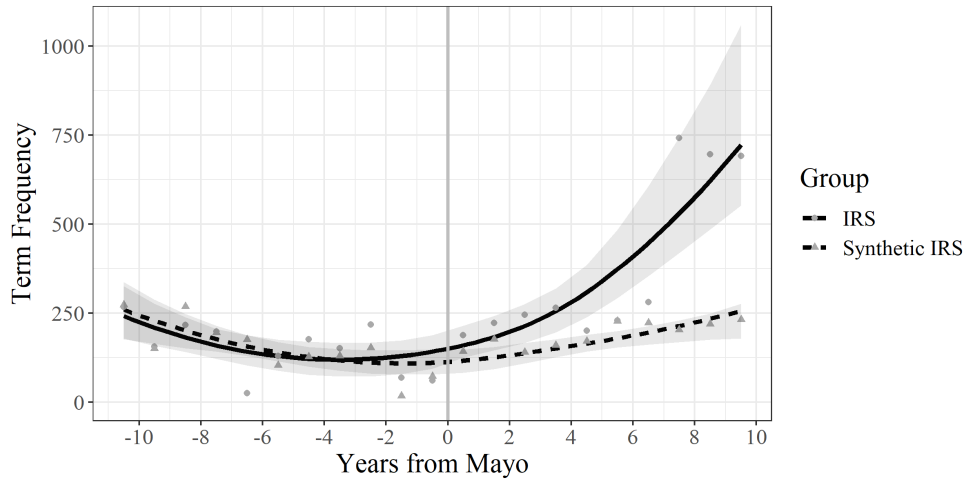


Figure 10: IRS v. Synthetic IRS, Statutory Term Frequency, 2000-2020

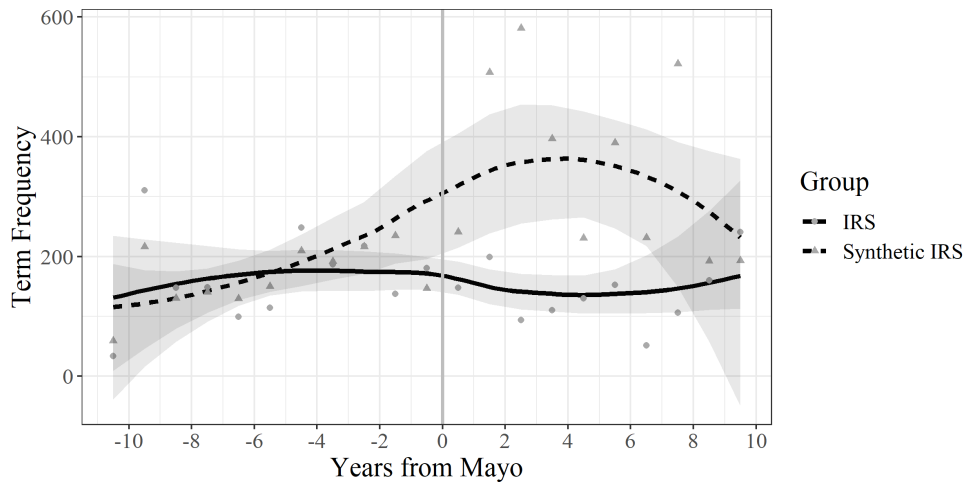


Figure 11: IRS v. Synthetic IRS, Preamble Length, 2000-2020

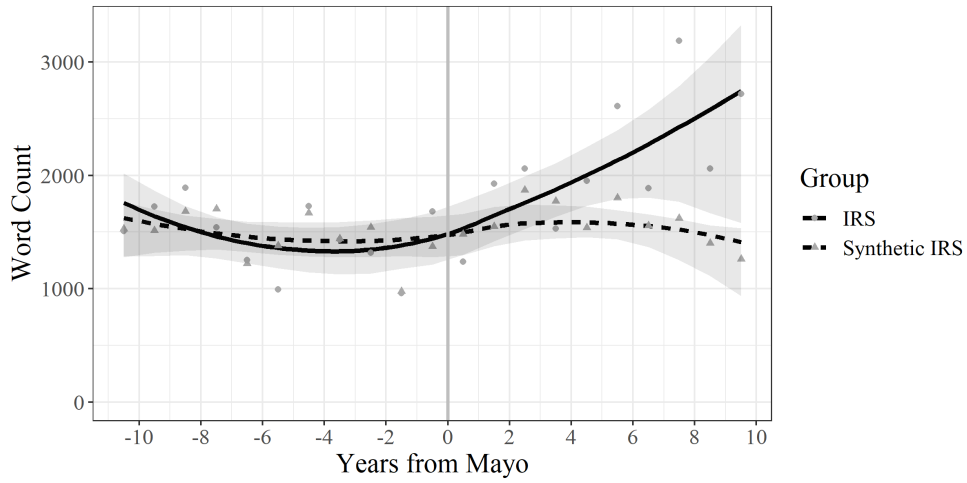
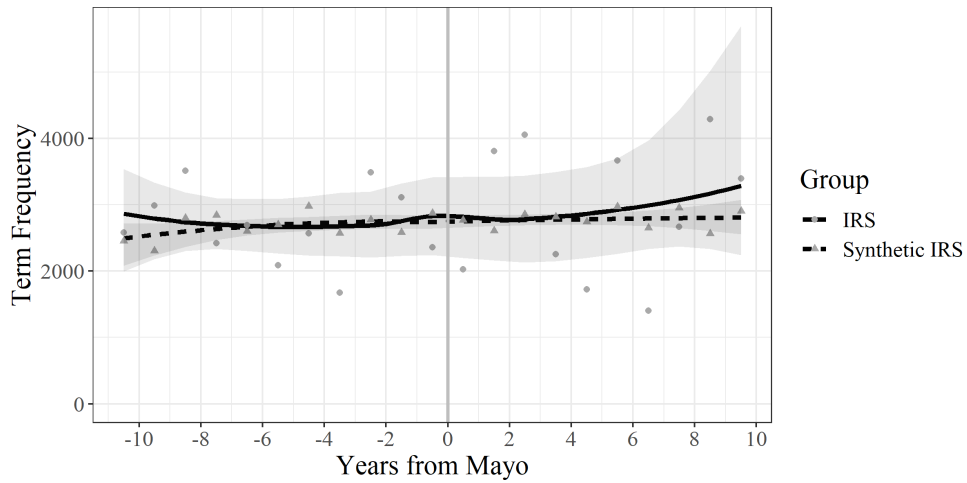


Figure 12: IRS v. Synthetic IRS, Commenter Term Frequency, 2000-2020



In addition to graphical results, average estimated effects can be calculated for the post-treatment period, and confidence intervals can be calculated by bootstrapping the results.¹⁹⁰

190. Specifically, the standard errors were calculated using non-parametric bootstrapping. See Xu & Liu, *supra* note 186.

Table 8: Synthetic Controls Estimates of Average Treatment Effects, 2000-2020

	Treatment Effect	95% Confidence Interval
Normative Term Frequency (per Million Words)	+186.7***	146.8 – 226.7
Statutory Term Frequency (per Million Words)	-209.5***	-352.9 – -66.1
Preamble Length	+531.2***	352.1 – 710.4
Commenter Term Frequency (per Million Words)	+144.0	-85.2 – 373.3

*** p<0.01, ** p<0.05, * p<0.1

Synthetic controls provide some additional assurance that the main results in this Article are valid. However, some doubt remains. Notably, the commenter term frequency results are small in magnitude and exceptionally noisy, yielding no statistically significant result. These initial results suggest that an effect may exist, but additional research is needed to confirm and especially to generalize these findings.

Conclusion

Although empirical scholars have extensively studied how courts apply *Chevron* deference, almost none have examined *Chevron's* effect on agencies. While agency administrators are aware of *Chevron* deference, many have suggested that *Chevron* may have little or no effect, either because it is rarely applied by courts or because its alternatives are equally deferential.

This Article exploits the unique shift from *National Muffler* deference to *Chevron* deference in 2011 to study the effect of *Chevron* deference on agency rulemaking. It suggests that preambles to regulations that received newly heightened deference increasingly discussed normative goals rather than underlying statutes. This in turn suggests that heightened *Chevron* deference encourages agencies to emphasize policymaking rather than legal interpretation.

In addition, the Article proposes and provides some empirical evidence for a new theory of the relationship between judicial deference and rulemaking procedures: that agencies view the requirements of *State Farm* and the APA as the price of deference and will therefore spend more effort complying with those requirements if granted more judicial deference. Empirical research suggests that, after *Mayo*, the IRS began to write longer preambles with more extensive discussion of public comments. However, these latter results are more tentative,

and additional research is warranted on the plausibility of the price-of-deference model.

These findings have a variety of implications for the study of administrative and tax law. They contribute to the literature on tax exceptionalism, suggesting that tax law's special deference regime restrained Treasury from interpreting tax statutes with the same emphasis on policy that has predominated at other agencies. *Mayo* therefore set the stage for the pronounced shift toward normative goals reflected in modern tax regulations. In turn, this shift affords each new presidential administration considerable freedom in reforming administrative tax law: for example, the Trump administration's sweeping directive that Treasury cultivate tax regulations that are "simple, fair, efficient, and pro-growth."¹⁹¹

Most broadly, the findings suggest that judicial deference does substantially affect agency rulemaking. This provides a practical counter to arguments that *Chevron* is indistinguishable from other deference regimes. Some might conclude from these findings that *Chevron* allows agencies to encroach on policymaking territory that properly belongs to Congress, while also producing wastefully overlong preambles that go unread except by privileged commenters. Others might conclude that *Chevron* empowers agencies to properly deploy their expertise and encourages them to provide necessary guidance that responds to public feedback. Rather than favoring the arguments of either side, this Article provides the terrain for arguments about *Chevron* to take place. Regardless of theoretical commitments, this Article emphasizes the importance of *Chevron* in shaping agency behavior.

191. Exec. Order No. 13,789, 82 Fed. Reg. 19,317 (Apr. 21, 2017); *see also* Eliminating Unnecessary Tax Regulations, 84 Fed. Reg. 9231, 9231 (Mar. 14, 2019) (eliminating 296 regulations that Treasury deemed "no longer necessary because they do not have any current or future applicability"); Exec. Order No. 13,777, 82 Fed. Reg. 12,285 (Feb. 24, 2017) (requiring agencies to undertake reforms intended to "lower regulatory burdens on the American people").

Appendix

A. Modeling the Price of Deference

Take a simple utility model for a hypothetical agency:

$$E(U) = b \cdot d \cdot P(x) - x \quad (1)$$

where:

- $E(U)$ = Agency's expected utility from rulemaking
- b = Benefit from agency rulemaking, which is lost if court overturns rule
- d = Probability that a regulation will be held a reasonable interpretation of the statute under the relevant deference regime
- x = Effort expended by the agency in rulemaking procedures, which will be higher when preambles are more detailed and responsive to public comments
- $P(x)$ = Probability that a regulation will be held to comply with the procedural requirements of *State Farm* and section 553(c) of the APA. The greater the effort that the agency expends in complying with rulemaking procedures, the higher the likelihood that it will be held to comply with *State Farm* and section 553(c)

The model assumes that $P(x)$ increases strictly monotonically above 0 (i.e., increasing the probability of success increases utility) and has a negative second derivative (i.e., there are decreasing returns to procedural effort).¹⁹² Utility is simply linear with respect to the probability that a regulation will be upheld.

Proof that $\frac{\partial x}{\partial d} > 0$:

The point at which utility is maximized is:

$$\frac{\partial}{\partial x} [E(U)] = 0 = \frac{\partial}{\partial x} [b \cdot d \cdot P(x) - x]$$

Pulling out linear multipliers and solving for the first derivative of x :

$$0 = b \cdot d \cdot \frac{\partial}{\partial x} [P(x)] - 1$$

192. See *supra* note 65 and accompanying text.

$$P'(x) = \frac{1}{b \cdot d}$$

By the implicit function theorem, we can totally differentiate this expression with respect to d , which gives:

$$\frac{\partial x}{\partial d} = \frac{-1}{b \cdot d^2 \cdot P''(x)}$$

By assumption, $P''(x) < 0$ (i.e., $P(x)$ has a negative second derivative). Similarly, $b > 0$ (the agency has a positive benefit from successfully promulgating a regulation) and $d > 0$ by assumption, since d represents a constant probability greater than zero that a regulation will be upheld on deference grounds. Therefore:

$$\frac{\partial x}{\partial d} = \frac{-1}{b \cdot d^2 \cdot P''(x)} > 0$$

In non-mathematical terms, any increase in the deference level d should be met with an increase in effort level x .

Note that this proof will apply for any function $P(x)$ with monotonically increasing and marginally declining utility. For example, the graphs in the body of this Article model $P(x)$ as the standard logistic function:

$$P(x) = \frac{1}{1 + e^{-x}} \tag{2}$$

But any other concave, monotonically increasing function could be used. It should also be noted that many plausible alternative models could generate opposite results. For example, under a “satisficing” model of agency utility, where agency administrators attempt to meet some threshold likelihood that a regulation will be upheld but are indifferent to increases above that threshold, *Chevron* deference might discourage additional effort by increasing the baseline likelihood that a regulation will be upheld above the satisficing threshold. Or, alternatively, if the probability that a regulation is upheld as a function of procedural effort is sigmoid for values of procedural effort greater than zero, then the proof above, which relies on monotonic decreases in the first derivative, will no longer work. For these reasons and others, the theoretical model presented in this Article is intended to suggest that the price-of-deference model is theoretically plausible and merits further empirical study, rather than as itself acting as a piece of evidence in favor of the price-of-deference model.

B. Data

This Article analyzes a new dataset developed using bulk data in XML format, downloaded from FederalRegister.Gov. The XML schema for these data is available online.¹⁹³

FederalRegister.Gov is an official government source for the rulemaking activity of federal agencies. It includes final regulations, as well as proposed regulations and temporary regulations. It also includes more procedural documents, like notices for public comment sessions on regulations. This Article analyzes all rulemaking documents, including proposed rulemaking, since proposed rulemaking contains important information about the agency's justification for the rule and reflects the agency's expectations regarding judicial deference. However, all purely procedural documents (like notices scheduling public comment sessions) were dropped from the analysis.¹⁹⁴

FederalRegister.Gov includes regulations from 2000 to the present. This provides a good range for the difference-in-differences study conducted by this Article, but regulations issued prior to 2000 could be useful for ancillary research questions—for example, a single-difference analysis of the effect of *Chevron* in 1984.

C. Terms Analyzed

The terms analyzed in this Article are largely drawn from prior empirical work by myself and others, as noted in greater detail below.¹⁹⁵ Using a consistent set of terms across articles reduces researcher degrees of freedom and is intended

193. U.S. Gov't Printing Office, User Guide Document: Federal Register XML Rendition (2009), https://www.govinfo.gov/bulkdata/FR/resources/FDsys_OFR-XML_User-Guide-v1.pdf [<https://perma.cc/DV3G-MJKZ>].

194. Any rule that lacked a preamble was also dropped.

195. Bruhl, *supra* note 105, at 30-31, 38-39, 41, 53 (listing and describing the use of search terms to assess judicial purposivism, textualism, and canon use); Nancy Staudt et al., *Judging Statutes: Interpretive Regimes*, 38 LOY. L.A. L. REV. 1909, 1933-35, 1940-42, 1950-51, 1956-59 (2005); Choi, *supra* note 36, at 419-24 (discussing the selection of statutory, normative, and purposivist terms). The terms were primarily drawn from my prior article, *An Empirical Study of Statutory Interpretation in Tax Law*, with appropriate modifications to generalize a set of terms specific to tax law to other agencies. I removed terms specifically referring to tax administration and removed references to the Congressional Budget Office. I also made modifications based on my random review of specific regulatory preambles. In particular, I removed “complexity” as a normative term, since a substantial minority of its occurrences discussed some aspect of the conduct being regulated rather than the complexity of the regulations themselves. *E.g.*, Federal Motor Vehicle Safety Standards; Theft Protection, 71 Fed. Reg. 17,752 (Apr. 7, 2006) (to be codified at 49 C.F.R. pt. 571); Funding and Fiscal Affairs, Loan Policies and Operations, and Funding Operations; Investment Management, 77 Fed. Reg. 66,361 (Nov. 5, 2012) (to be codified at 12 C.F.R. pt. 615); Basel III Conforming Amendments Related to Cross-References, Subordinated Debt and Limits Based on Regulatory Capital, 79 Fed. Reg. 11,300 (Feb. 28, 2014) (to be codified at 12 C.F.R. pts. 1, 4, 5, 16, 23, 24, 28, 32, 34, 46, 116, 143, 145, 159, 160, 161, 163, 192). I also removed “committee report” as a legislative history term, as it led to false positives when referring to reports of non-congressional committees. *E.g.*, Irish Potatoes Grown in Washington; Modification of Special Purpose Shipment Regulations, 70 Fed. Reg. 44,252 (Aug. 2, 2005) (to be codified at 7 C.F.R. pt. 946); Apricots Grown in Designated Counties in Washington; Decreased Assessment Rate, 83 Fed. Reg. 4412 (Jan. 31, 2018) (to be codified at 7 C.F.R. pt. 922).

to allay concerns about cherry-picking terms to increase the statistical significance of results.¹⁹⁶ For purposes of calculating term frequencies, the terms are not case-sensitive and are therefore all listed in lower-case. In addition, prior to conducting the analysis, the text was cleaned and regularized by removing special characters and extra whitespace. Text matches included partial word matches—for example, the word “unfair” also includes “unfairness” and “unfairly.”¹⁹⁷

1. Normative Terms

Any occurrences of normative terms in sentences that also included terms associated with statutory interpretation (legislative history, textualist interpretive tools, or canons of construction) were excluded, in order to avoid misclassifying discussions of policy judgments that occur in the statutory process.¹⁹⁸

good public policy
public policy goal
public policy grounds
efficient administration
efficient enforcement
compliance burden
financial burden
administrative burden
regulatory burden

burdensome
compliance cost
complexity
intrusive
fairness
unfair
injustice
unjust
clarity

2. Statutory Terms

Unlike the other terms in this Article, a document’s statutory term frequency was determined based on the number of statutory *sentences*. A sentence was designated as statutory if it included at least one word from the column on the left below, and one word from the column on the right below. In addition, sentences that include the word “section” but also include the phrase “executive order” were dropped in order to avoid false positives discussing the interpretation of executive orders rather than statutes.¹⁹⁹

196. Joseph P. Simmons, Leif D. Nelson & Uri Simonsohn, False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant, 22 PSYCH. SCI. 1359, 1359 (2011) (describing researcher degrees of freedom as “exploratory behavior” by researchers who tend to “explore various analytic alternatives, to search for a combination that yields ‘statistical significance,’ and to then report only what ‘worked’”).

197. This refers only to differences in prefixes or suffixes. In contrast, “financially burdensome” would not be a match for “financial burden” (although it would separately be a match for “burdensome”).

198. See Choi, *supra* note 36, at 423 (applying the same exclusion).

199. *E.g.*, Reportable Quantity Adjustment for Isophorone Diisocyanate, 71 Fed. Reg. 53,331 (Sept. 11, 2006) (to be codified at 40 C.F.R. pt. 355); Regulation of Fuels and Fuel Additives: Modifications to Reformulated Gasoline Covered Area Provisions, 67 Fed. Reg. 38,398 (June 4, 2002) (to be codified at 40 C.F.R. pt. 80).

Legal Analysis, Policy Analysis, and the Price of Deference

Note that the terms selected require relatively intense engagement with legislation for a sentence to be considered an act of statutory interpretation. Most regulations will trivially mention some statute; the terms selected below aim to restrict positive results to serious attempts to interpret the statute, rather than casual references.

Includes:	AND	Includes:
<i>construe</i>		<i>code</i>
<i>construing</i>		<i>congress</i>
<i>construction</i>		<i>legislation</i>
<i>interpret</i> ²⁰⁰		<i>pub. l. no.</i>
<i>read</i>		<i>section</i>
		<i>statute</i>
		<i>statutory</i>
		<i>u.s.c.</i>

3. Legislative History

Congressional Reports

<i>conference report</i>	<i>h.r. rept.</i>
<i>conf. rep.</i>	<i>h. r. rept.</i>
<i>conf. rpt.</i>	<i>h.r.rep.</i>
<i>conf. rept.</i>	<i>h.r.rpt.</i>
<i>conf.rpt.</i>	<i>h.r.rept.</i>
<i>conf.rept.</i>	<i>senate report</i>
<i>house report</i>	<i>s. rep.</i>
<i>h. rep.</i>	<i>s. rpt.</i>
<i>h. rpt.</i>	<i>s. rept.</i>
<i>h. rept.</i>	<i>s.rep.</i>
<i>h.rep.</i>	<i>s.rpt.</i>
<i>h.rpt.</i>	<i>s.rept.</i>
<i>h.rept.</i>	<i>comm. rep.</i>
<i>h.r. rep.</i>	<i>comm. rpt.</i>
<i>h. r. rept.</i>	<i>comm. rept.</i>
<i>h.r. rpt.</i>	<i>comm.rep.</i>
<i>h. r. rpt.</i>	<i>comm.rpt.</i>
	<i>comm.rept.</i>

Congressional Hearings

200. The term “interpretative” is excluded from this count, although it contains “interpret” as a substring.

congressional hearing
congressional record
cong. rec.
cong.rec.
rec. doc.

committee hearing
senate hearing
house hearing
conference hearing

Miscellaneous Legislative History

legislative history
history of the legislation
conference committee
senate committee
*s. comm.*²⁰¹

s. subcomm.
house committee
h.r. comm.
h. subcomm.
h. r. subcomm.

4. Commenters

Because the term “comment” returned many false positives, a more limited list of terms referring specifically to commenters themselves was used.

commentator
commenter

commentor

D. Distinguishing Interpretative and Legislative Regulations

The line between interpretative and legislative regulations has changed over time, largely as a result of *Mayo*. Prior to *Mayo*, it was widely believed (including by the IRS) that legislative regulations were promulgated under specific grants of regulatory authority, whereas interpretative regulations followed from Treasury’s general authority to promulgate regulations under section 7805(a) of the Code.²⁰² The IRS explicitly disavowed this view after *Mayo*,²⁰³ instead applying the definition prevalent in administrative law outside of tax, that legislative regulations demand notice and comment whereas interpretative ones do not.²⁰⁴ But even before *Mayo*, status as an interpretative regulation could be unclear. The IRS frequently (both before and after *Mayo*) cited section 7805 as the sole

201. To avoid false positives, the text was pre-processed to exclude the term “pers. comm.,” an abbreviation for “personal communication.” *E.g.*, Endangered and Threatened Wildlife and Plants; Final Rule to List the Vermilion Darter as Endangered, 66 Fed. Reg. 59,367 (Nov. 28, 2001) (to be codified at 50 C.F.R. pt. 17).

202. Hickman, *supra* note 88, at 1545 (“Since long before *Chevron*, however, and consistent with the tax community’s categorization, the Treasury has taken the position that its general authority regulations are interpretative only . . .”).

203. I.R.S., *supra* note 88, § 2.1.1.2.8.1 (“Whether a regulation is promulgated under a specific grant of authority in the Internal Revenue Code does not govern whether the regulation is interpretative or legislative.”).

204. I.R.S., INTERNAL REVENUE MANUAL, § 32.1.1.2.6.1 (Sept. 23, 2011) (“The Administrative Procedure Act (APA) exempts interpretative rules from the APA’s notice and comment requirements.”).

source of authority for its regulations, even when a more specific grant of regulatory authority existed.²⁰⁵ And after *Mayo*, internal IRS policy generally directs employees to conduct notice and comment for all Treasury regulations, including interpretative regulations.²⁰⁶

Despite occasional fuzziness, the best line to separate legislative from interpretative regulations under the pre-*Mayo* understanding of these terms remains whether their authority derives from section 7805, or some other section of the Code.²⁰⁷ Accordingly, this Article categorizes Treasury regulations according to their underlying authority. This is relatively straightforward, because each Treasury regulation (excluding amendments and corrections) explicitly states the authority on which it relies.²⁰⁸ The authority for each regulation is cited in the Federal Register, as required by the APA.²⁰⁹

For this Article, I programmed an algorithm to identify which regulations cited section 7805 as their sole authority (classifying these as interpretative regulations), cited sections of the Code other than 7805 as their sole authority (classifying these as legislative regulations), or cited both section 7805 and some other section of the Code (generally excluding these from the analysis). Only complete regulations were counted for purposes of this Article, so that amendments and corrections that did not cite any authority were excluded. Although the Federal Register XML data were of high quality, I individually reviewed each of the authorities cited in order to avoid any false positives or negatives. For purposes of the regression analysis, Treasury regulations that were entirely legislative were considered “other” (non-interpretative) regulations and were grouped with non-Treasury regulations, but Treasury regulations that cited both specific and general authority were dropped from the regression analysis altogether. Running the same regressions without dropping these specific-and-general-authority regulations produces essentially identical results, with the coefficients on the variable of interest in each regression taking the same sign and statistical significance, and with the point estimates falling within 5% of those shown above.

Table 9 summarizes the number of Treasury regulations of each type:

205. Hickman, *supra* note 88, at 1544 n.27 (citing many cases where Treasury was explicitly directed to make rules regarding certain sections of the Code, but nonetheless cited section 7805(a) as the sole authority in its rulemaking). One former IRS official recounted to me that, prior to *Mayo*, he and the IRS considered essentially all regulations interpretative.

206. DEPT. OF TREASURY, POLICY STATEMENT ON THE TAX REGULATORY PROCESS 1 (2019), <https://home.treasury.gov/system/files/131/Policy-Statement-on-the-Tax-Regulatory-Process.pdf>.

207. See *supra* Section I.D; Hickman, *supra* note 88, at 1557 (“Although the practical difference is not always apparent, in [jurisdictions that accorded some Treasury regulations only *National Muffler* deference, rather than *Chevron* deference], specific authority regulations are given ‘controlling weight’ pursuant to *Chevron* while general authority regulations promulgated under I.R.C. § 7805(a) are given only ‘considerable weight’ under *National Muffler*.”).

208. See *infra* Appendix Section D (describing in greater detail the process the Article uses to separate interpretative and legislative regulations).

209. 5 U.S.C. § 553(b)(2) (2018).

Table 9: Number of Treasury Regulations by Type, 2000-2020

Treasury Regulation Type	<i>N</i>
All Tax Regulations	1480
Interpretative Tax Regulations	1096
Legislative Tax Regulations	20
Other Tax Regulations (citing both section 7805 and some other section of the Code)	381

Ideally, the IRS would conduct notice and comment by issuing proposed regulations and then inviting comments on those proposed regulations, before issuing final regulations that incorporate changes in response to comments. In reality, the IRS often feels obliged to issue temporary regulations with immediate effect and then subsequently requests comments on those temporary regulations. The IRS claims that these temporary regulations fall within the “good cause” exception of the APA.²¹⁰ Some critics have argued that Treasury thereby fails to comply with the APA and that many of these temporary regulations are invalid.²¹¹ Nevertheless, because the IRS continues to maintain that its procedures regarding temporary regulations comply with the APA, and therefore that these regulations receive *Chevron* deference like any others, temporary regulations are not treated differently from other regulations for purposes of this Article.

Another nuance is that Treasury declines to open a small minority of regulations to notice and comment altogether.²¹² These regulations are generally exempt from notice and comment requirements under the APA’s good cause exception or its procedural rule exception.²¹³ Because these regulations are rare and hard to identify at scale, and because Treasury’s beliefs regarding whether these regulations should receive *Chevron* deference are unclear (in many cases, they amend or remove existing regulations, making it difficult to apply or not apply *Chevron* to them independently), they are not removed from the sample for purposes of this Article. To the extent that these regulations are not eligible for *Chevron* deference, including them in the sample will tend to attenuate the measured effect of *Chevron*, implying that point estimates will be too small in magnitude.

210. DEPT. OF TREASURY, *supra* note 206, at 1.

211. *Supra* note 137.

212. Hickman, *supra* note 93, at 1749 (finding that 11 out of 232 regulations from a three-year sample were neither preceded nor followed by notice and comment).

213. *Id.* at 1789 (“In sum, most of the projects in which Treasury issued final regulations without the benefit of notice and comment seem to be good candidates for either the procedural rule exception or the good cause exception.”). Of the eleven in Hickman’s sample, Treasury (incorrectly) attempted to argue that one did not require notice and comment because it was an interpretative rule. *Id.* at 1787.

The conspicuous shortage of pure legislative tax regulations—those that solely draw their authority from some section of the tax code other than section 7805—precludes another potential robustness check, namely analysis that solely considers tax regulations, using interpretative tax regulations as the treatment and legislative tax regulations as the control group. Because pure legislative tax regulations are so rare (18 in this sample), *N* is too small to conduct a meaningful analysis.

A final potential issue is Treasury’s unusual format for citing statutory authority. Many Treasury regulations that cite only Section 7805 of the Code (making them *prima facie* interpretative) specifically cite “Section 7805 * * *.”²¹⁴ According to one source, the three asterisks are “an agreed-upon signal from the I.R.S. to the Office of the Federal Register (O.F.R.)” to incorporate by reference specific authority cited at the beginning of the relevant volume of the Code of Federal Regulations (CFR). As a result, if the relevant volume of the CFR cited specific authority, a regulation citing “7805 * * *” could arguably be legislative, rather than interpretative.²¹⁵ If so, many of the regulations classified as interpretative by this Article might actually be legislative. If legislative tax regulations underwent no change in deference status before and after *Mayo* (because they have always received *Chevron* deference), then including those covertly legislative regulations in the category of interpretative tax regulations could introduce additional noise in the regression, attenuating the measured effect of *Mayo* on truly interpretative tax regulations. As a result, point estimates for the magnitude of *Mayo*’s effects may be underestimated, and the true effect of *Mayo* and *Chevron* might actually be greater.²¹⁶

One method to test whether the inclusion of three-asterisk regulations has biased the results is simply to conduct the regressions, excluding three-asterisk regulations from the treatment group. In these regressions, the main results presented in this Article—the changes in normative terms, statutory terms, preamble length, and references to commenters—remained statistically significant and in most cases increased in magnitude, *except* that the first-step regression regarding statutory terms lost statistical significance (although it remained negative). In general, these results suggest that Treasury may have

214. These asterisks generally appear when a regulatory document amends an existing regulation. *E.g.*, The Solely for Voting Stock Requirement in Certain Corporate Reorganizations, 65 Fed. Reg. 31,805, 31,806 (May 19, 2000) (to be codified at 26 C.F.R. pt. 1); Section 6038—Returns Required With Respect to Controlled Foreign Partnerships, 67 Fed. Reg. 78,174, 78,175 (Dec. 23, 2002) (to be codified at 26 C.F.R. pts. 1, 602).

215. Hickman, *supra* note 86, at 1753 n.99 (noting that she has received this comment, but adding that the supposed use of the three asterisks to indicate specific authority is “inherently cryptic” and arguing that it does not satisfy the authority citation requirement of section 553(b)(2) of the APA). *See also* I.R.S., INTERNAL REVENUE MANUAL, § 32.1.5.7.4.2, exs. 1, 2 (Aug. 11, 2004) (arguably following this custom, albeit without explanation).

216. It is also theoretically possible that the inclusion of the three-asterisk regulations introduces some sort of omitted variable bias that causes the sign of the effect to be misestimated (e.g., to estimate the causal effect on normative term frequency as positive rather than negative). As noted further below, this is not consistent with the data, and moreover it is difficult to imagine why this would be so.

expected *Chevron* deference for at least a subset of the three-asterisk regulations and that the results in this Article, while directionally correct, may actually underestimate effect size.

Nevertheless, this restricted regression is less reliable because of the relatively small number of tax regulations that solely cite section 7805 without including asterisks (30 out of 1084), as well as lingering uncertainty about the formal meaning of the asterisks and the extent to which this practice complied with the APA.²¹⁷ Even assuming that the IRS's informal signal to the OFR was internally understood and did comply with the APA, it is unclear which authority cited in the relevant CFR volume is intended to be picked up by the asterisks. It appears that many regulations simply cited "Section 7805 * * *" as a matter of course, making it impossible to determine which were intended to be legislative and which were intended to be interpretative. Consequently, this Article still considers three-asterisk regulations to be interpretative for purposes of the main analysis, but this issue provides additional reason to treat the point estimates in this Article as lower bounds in magnitude.

E. Descriptive Statistics

The dataset used in this Article includes 467 separate agencies. The number of regulations per agency is not normally distributed—the most prolific regulator, the Federal Aviation Administration (FAA), produced 15,262 regulations during the time frame studied, around 22% of the total. Additional descriptive statistics regarding the data and variables are below.

217. See *supra* note 215.

Table 10: Descriptive Statistics on Variables

Variables	(1) N^{218}	(2) Mean	(3) Median	(4) Min	(5) Max	(6) σ
Normative Term Frequency (per Million Words)	69,870	185.8	0	0	35,714	621.5
Statutory Term Frequency (per Million Words)	69,870	107.7	0	0	40,000	703.0
Legislative History Term Frequency (per Million Words)	69,870	10.68	0	0	16,129	151.9
Commenter Term Frequency (per Million Words)	69,870	599.0	0	0	36,145	1910
Preamble Word Count	69,956	1352	765	0	35,245	1860
Regulation Word Count	69,956	735.8	225	0	34,822	1588
Reading Level	69,956	13.2	13.4	-22.21	61.55	2.138
Democratic Administration	69,956	0.384	0	0	1	0.486

Table 11 lists the number of regulations in the dataset per year, as well as a breakdown of the number of regulations issued by each of the ten most prolific agencies, which includes the IRS. In order, these agencies are: the FAA, the Environmental Protection Agency (EPA), the Coast Guard, the National Oceanic and Atmospheric Administration (NOAA), the Federal Communications Commission (FCC), the Food and Drug Administration (FDA), the Agricultural Marketing Service (AMS), the Federal Emergency Management Agency (FEMA), the IRS, and the Animal and Plant Health Inspection Service (APHIS).

Note that the number of regulations issued by agencies may not capture every dimension of the agency’s activities. For example, while the FAA issued over thirteen times more regulations than the IRS in the period studied in this Article, FAA preambles are generally shorter, and the word count of those preambles was only about six times greater at the FAA than the IRS for the same period.²¹⁹

218. N varies because a small number of regulations lack preambles, making it impossible to calculate term frequency (because the term frequency would be zero divided by zero). These observations are dropped from the regression analysis when term frequency is used.

219. FAA preambles had a total of 12,133,693 words, versus 2,018,494 for the IRS.

Table 11: Number of Regulations for Top 10 Agencies by Regulations Promulgated, 2000-2020

Year	All	(1) FAA	(2) EPA	(3) Coast Guard	(4) NOAA	(5) FCC	(6) FDA	(7) AMS	(8) FEMA	(9) IRS	(10) APHIS
2000	3708	871	488	224	267	331	188	98	44	40	78
2001	3597	783	596	364	275	309	120	65	51	45	61
2002	3682	703	526	405	269	302	133	103	68	54	90
2003	3731	861	481	353	254	268	153	94	83	66	81
2004	3630	905	456	328	259	277	163	85	81	65	75
2005	3518	896	471	301	279	222	112	92	68	71	59
2006	3355	739	458	295	298	186	127	91	69	89	80
2007	3222	755	431	311	288	101	130	87	78	79	71
2008	3431	785	462	285	322	155	108	72	89	81	46
2009	3105	632	408	305	279	126	95	63	72	46	50
2010	3200	711	428	377	302	95	86	60	98	40	33
2011	3372	699	461	445	284	122	90	65	89	61	43
2012	3298	665	580	531	278	104	73	42	73	45	26
2013	3263	763	476	504	258	124	68	70	54	42	29
2014	3206	629	495	482	287	129	87	63	52	60	32
2015	2987	560	514	447	276	79	78	44	35	44	27
2016	3363	683	500	429	301	79	116	60	30	61	21
2017	3086	701	550	507	279	102	87	35	25	21	15
2018	3153	720	482	556	283	120	86	50	28	15	19
2019	2713	507	446	392	265	85	47	52	23	43	7
2020	2955	694	412	201	294	114	39	41	27	48	11
Total	69,575	15,262	10,121	8042	5897	3430	2186	1432	1237	1116	954

F. Non-Normal Distribution of Term Frequencies

The most common form of regression analysis, ordinary least squares (OLS) regression, assumes that regression residuals are homoskedastic and normally distributed.²²⁰ However, these assumptions are implausible for the data analyzed in this Article. The term frequencies described in this Article are (1) *semicontinuous*, meaning that they cannot be less than zero with respect to any text, but can vary continuously above zero; (2) *zero-inflated*, meaning that there is a disproportionate quantity of texts that have a term frequency of zero; and (3) *log-normal*, meaning that they do not follow a normal distribution even for term frequencies above zero, although they are normally distributed when log-

220. Jacob Cohen et al., *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* 480 (3d ed. 2003).

transformed.²²¹ All these features make the assumptions of normally distributed and homoskedastic residuals unlikely.²²²

Table 10 shows the problem of zero-inflation by noting that for each of the term frequencies studied, the median value is zero. The problem can also be seen in the histograms below, in Figure 13:

Figure 13: Histograms of Dependent Variables, 2000-2020

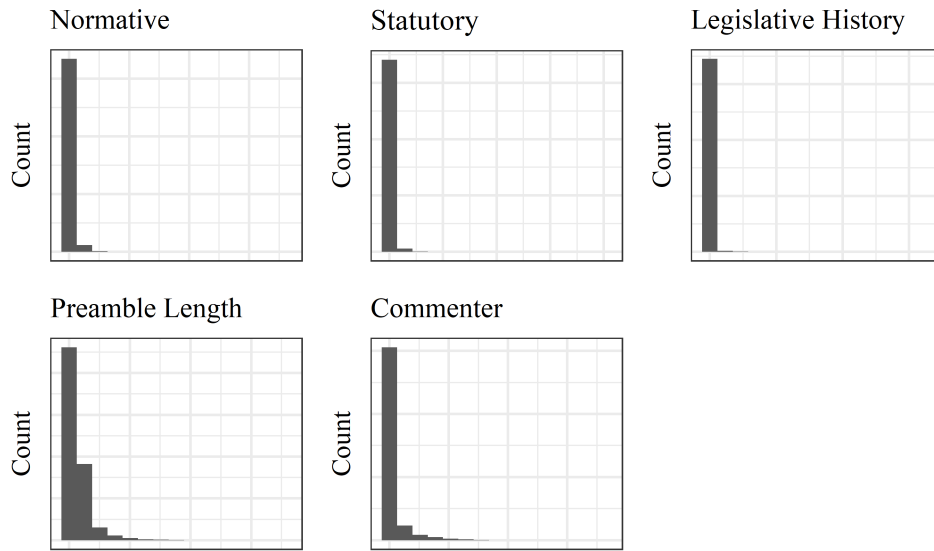


Figure 13 shows that most of the data are significantly zero-inflated—a substantial majority of the regulatory preambles have zero uses of statutory, normative, legislative history, and commenter terms, respectively. (The data are not zero-inflated with respect to preamble length, however.²²³) In addition, the nonzero term frequencies are also not normally distributed—they are log-normal.

The log-normality of the nonzero term frequencies can be demonstrated by log-transforming them as follows for each term frequency y :

$$\tilde{y} = \log (y) \tag{3}$$

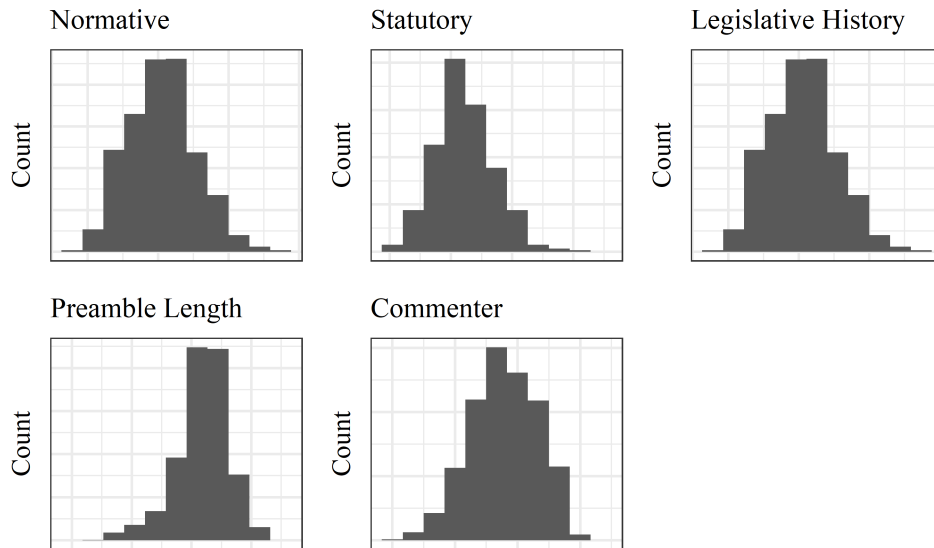
221. Preamble length is a count variable, so it is not semicontinuous or zero-inflated, but it is log-normal, as described below, and it cannot be less than zero.

222. While these distributional features describe the dependent variable and not the residuals, they affect the distribution of residuals as well—for example, a hard cutoff at zero means that the distribution of residuals is likely to have a left cutoff with a long right tail.

223. Unsurprisingly, very few regulations are promulgated without a preamble—86 out of 69,956 in the sample.

After log-transformation, all of the regression dependent variables become approximately normally distributed, as Figure 14 shows. This motivates the choice of the log link discussed in Section H of the Appendix.

Figure 14: Logged Histograms of Dependent Variables, 2000-2020



As the following Section describes, two-part regression addresses issues of semi-continuity, zero-inflation, and log-normality, making it the appropriate choice of model for the term frequencies in this Article. The preamble length regressions use a generalized linear model (GLM) that is similarly adapted to log-normally distributed count data.

G. Regression Models: OLS, GLM, and Two-Part Regression

This Article implements a difference-in-differences study through regression analysis. It takes each rule promulgated between 2000 and 2020 as a separate observation. In addition to controls, it analyzes three variables of special interest: (1) a dummy variable coded as 1 if the regulation was promulgated after *Mayo* (regardless of agency), and 0 otherwise; (2) a dummy variable coded as 1 if the regulation was an interpretative tax regulation, and 0 otherwise; and (3) an interaction variable, generated by multiplying the prior two variables. Five different dependent variables are studied through the regressions: the term frequency of statutory, normative, legislative history, and commenter terms, and the length of regulatory preambles.

The coefficient generated in the regression for the post-*Mayo* dummy represents the change after *Mayo* for regulations other than interpretative tax regulations. The coefficient for the interpretative tax regulation dummy represents the difference between interpretative tax regulations and other

regulations prior to *Mayo*. And the key variable of interest for purposes of our study is the interaction term. The coefficient on the interaction term represents the *additional* marginal effect that *Mayo* had specifically on interpretative tax regulations. The interaction term is therefore the treatment effect or difference in differences—the difference between interpretative tax regulations and other regulations, in the differences before and after *Mayo*.

In addition to the dummy variables and interaction term, subsequent regressions also include control variables for time trends (piecewise, pre- and post-*Mayo*), length of the regulation associated with the preamble, a dummy variable indicating whether the preamble has a high reading level or a low reading level,²²⁴ a dummy reflecting the party of the sitting president when the regulation was issued, and dummy variables identifying the type of regulation (rule, notice, or other²²⁵). The basic OLS model is as follows:

$$y_i = \beta_0 + \beta_1 \cdot PostMayo_i + \beta_2 \cdot Interpretative\ Tax\ Reg_i + \beta_3 \cdot PostMayo_i \cdot Interpretative\ Tax\ Reg_i + \epsilon_i \quad (4)$$

Controlling for the time trend (separately estimated for the periods before and after *Mayo*), the model is:

$$y_i = \beta_0 + \beta_1 \cdot PostMayo_i + \beta_2 \cdot Interpretative\ Tax\ Reg_i + \beta_3 \cdot PostMayo_i \cdot Interpretative\ Tax\ Reg_i + \beta_4 \cdot Year_i + \beta_5 \cdot Year_i \cdot PostMayo_i + \epsilon_i \quad (5)$$

Finally, with full controls, the model is:

$$y_i = \beta_0 + \beta_1 \cdot PostMayo_i + \beta_2 \cdot Interpretative\ Tax\ Reg_i + \beta_3 \cdot PostMayo_i \cdot Interpretative\ Tax\ Reg_i + \beta_4 \cdot Year_i + \beta_5 \cdot Year_i \cdot PostMayo_i + \beta_6 \cdot Regulation\ Length_i + \beta_7 \cdot High\ Reading\ Level_i + \beta_8 \cdot Democratic\ Administration_i + \beta_9 \cdot Rule_i + \beta_{10} \cdot Notice_i + \epsilon_i \quad (6)$$

However, OLS models are problematic for the dataset in this Article, for the reasons discussed above. Consequently, I instead primarily rely on a GLM (for the preamble length regression) or a two-part regression model (for the term frequency regressions). The two-part model has become standard in a variety of

224. A preamble is designated as having a high reading level if its Coleman-Liau reading level is greater than 12, meaning reading appropriate for someone with a high school degree or more. The main benefit of controlling for reading level is that it controls for regulations with trivially short preambles, which have very low (and sometimes negative) reading levels. Thus, the use of a dummy variable, which captures the variation of interest in reading level despite relatively low variance in reading level among preambles (the standard deviation is 2.14).

225. “Other” documents are not explicitly controlled for, but they form the baseline when a document is neither a rule nor a notice.

fields with semicontinuous data that present zero-inflation as well as log-normality.²²⁶

A two-part regression model separates estimates of term frequency into two parts. First, a logit or probit model estimates the likelihood that term frequency is nonzero.²²⁷ Second, an OLS model or GLM estimates term frequency magnitude conditional on term frequency being nonzero. Intuitively, one could imagine an agency making an initial decision on whether to conduct statutory (or normative, or legislative history) analysis at all, and then making a second separate decision about how much to discuss the statute (or normative concerns, or legislative history).

For this Article, I use the two-part model developed by Naihua Duan et al.²²⁸ and implemented by Federico Belotti et al.²²⁹ For the first part, I specify a basic logit model as follows²³⁰:

$$\begin{aligned} \text{logit}[P(tf_i = 0)] &= \beta_0^1 + \beta_1^1 \cdot \text{PostMayo}_i + \beta_2^1 \cdot \\ &\text{Interpretative Tax Reg}_i + \beta_3^1 \cdot \text{PostMayo}_i \cdot \\ &\text{Interpretative Tax Reg}_i + \epsilon_i \end{aligned} \quad (7)$$

The second part of the regression uses a GLM. A GLM is a generalized version of an OLS model that relaxes some of the assumptions used in OLS regression, specifically the assumptions of linear functional form and homoskedasticity.²³¹ The basic GLM is as follows (noting that references to “log” in this Article are to the natural logarithm, sometimes also denoted “ln”)²³²:

$$\begin{aligned} \log[y_i | y_i > 0] &= \beta_0^2 + \beta_1^2 \cdot \text{PostMayo}_i + \beta_2^2 \cdot \\ &\text{Interpretative Tax Reg}_i + \beta_3^2 \cdot \text{PostMayo}_i \cdot \\ &\text{Interpretative Tax Reg}_i + \epsilon_i \end{aligned} \quad (8)$$

226. Partha Deb & Edward C. Norton, *Modeling Health Care Expenditures and Use*, 39 ANN. REV. PUB. HEALTH 489, 495 (2018) (“The health econometrics literature has settled on the two-part model as the best way to model a dependent variable with a large mass at zero and many positive values.”). For early examples of the two-part model, see generally J.A. Cole & J.D.F. Sherriff, *Some Single- and Multi-Site Models of Rainfall Within Discrete Time Increments*, 17 J. HYDROLOGY 97 (1972) (applying an early version of a two-part regression model to estimate rainfall); Naihua Duan et al., *Choosing Between the Sample-Selection Model and the Multi-Part Model*, 2 J. BUS. & ECON. STATS. 283 (1984) (applying a two-part model to estimate healthcare expenditures).

227. A probit model may also be used. *Id.* at 496 (“This choice is generally innocuous in that there is never a substantial difference between logit and probit.”).

228. Duan et al., *supra* note 226.

229. Federico Belotti et al., *Twopm: Two-Part Models*, 15 STATA J. 3 (2015).

230. See Yongi Min & Alan Agresti, *Random Effect Models for Repeated Measures of Zero-Inflated Count Data*, 5 STATISTICAL MODELLING. 1, 11 (specifying a similar model).

231. Deb & Norton, *supra* note 226, at 494-95 (“The GLM generalizes the ordinary linear regression model by allowing the expectation of the outcome variable to be a function (known as the link function) of the linear index of covariates, not simply a linear function of the index. Expenditure data, for example, often fit best with a log link, meaning that the natural logarithm of the expected value of the dependent variable is modeled as the linear index In addition, GLMs also explicitly model the heteroskedasticity. GLMs allow the variance of the outcome to be a function of its predicted value by the choice of an appropriate distribution family.”).

232. See Min & Agresti, *supra* note 230, at 11.

The two models, including time trend controls, are:

$$\begin{aligned} \text{logit}[P(y_i = 0)] &= \beta_0^1 + \beta_1^1 \cdot \text{PostMayo}_i + \beta_2^1 \cdot \\ &\text{Interpretative Tax Reg}_i + \beta_3^1 \cdot \text{PostMayo}_i \cdot \\ &\text{Interpretative Tax Reg}_i + \beta_4^1 \cdot \text{Year}_i + \beta_5^1 \cdot \text{Year}_i \cdot \text{PostMayo}_i + \epsilon_i \end{aligned} \quad (9)$$

$$\begin{aligned} \text{log}[y_i | y_i > 0] &= \beta_0^2 + \beta_1^2 \cdot \text{PostMayo}_i + \\ &\beta_2^2 \cdot \text{Interpretative Tax Reg}_i + \beta_3^2 \cdot \text{PostMayo}_i \cdot \\ &\text{Interpretative Tax Reg}_i + \beta_4^2 \cdot \text{Year}_i + \beta_5^2 \cdot \text{Year}_i \cdot \text{PostMayo}_i + \epsilon_i \end{aligned} \quad (10)$$

Finally, including full controls, the models are:

$$\begin{aligned} \text{logit}[P(y_i = 0)] &= \beta_0^1 + \beta_1^1 \cdot \text{PostMayo}_i + \beta_2^1 \cdot \\ &\text{Interpretative Tax Reg}_i + \beta_3^1 \cdot \text{PostMayo}_i \cdot \\ &\text{Interpretative Tax Reg}_i + \beta_4^1 \cdot \text{Year}_i + \beta_5^1 \cdot \text{Year}_i \cdot \text{PostMayo}_i + \\ &\beta_6^1 \cdot \text{Regulation Length}_i + \beta_7^1 \cdot \text{High Reading Level}_i + \\ &\beta_8^1 \cdot \text{Democratic Administration}_i + \beta_9^1 \cdot \text{Rule}_i + \beta_{10}^1 \cdot \text{Notice}_i + \epsilon_i \end{aligned} \quad (11)$$

$$\begin{aligned} \text{log}[y_i | y_i > 0] &= \beta_0^2 + \beta_1^2 \cdot \text{PostMayo}_i + \beta_2^2 \cdot \\ &\text{Interpretative Tax Reg}_i + \beta_3^2 \cdot \text{PostMayo}_i \cdot \\ &\text{Interpretative Tax Reg}_i + \beta_4^2 \cdot \text{Year}_i + \beta_5^2 \cdot \text{Year}_i \cdot \\ &\text{PostMayo}_i + \beta_6^2 \cdot \text{Regulation Length}_i + \\ &\beta_7^2 \cdot \text{High Reading Level}_i + \beta_8^2 \cdot \\ &\text{Democratic Administration}_i + \beta_9^2 \cdot \text{Rule}_i + \beta_{10}^2 \cdot \text{Notice}_i + \epsilon_i \end{aligned} \quad (12)$$

In addition to the separate estimates of the marginal effect of each independent variable in the first and second part of the two-part regression, we can combine the two marginal effects to calculate the overall average marginal effect of each independent variable on the dependent variable. The formula for this calculation, of the effect of any independent variable x_i on dependent variable y_i , is²³³:

$$y_i = \hat{y}_i | x_i = (\hat{p}_i | x_i) \cdot (\hat{y}_i | y_i > 0, x_i) \quad (13)$$

Two-part regression is used for every regression except for the regression using preamble length as its dependent variable, because preamble length in the sample is log-normal but not zero-inflated. Consequently, a simple GLM regression is used for preamble length.

Because the absolute values of changes are difficult to interpret without understanding average values before and after, this Article also presents semi-elasticities, reflecting the percentage change in the dependent variable given a one-unit change in a relevant independent variable. Modifying Equation 13 to

233. Belotti et al., *supra* note 229, at 7.

calculate semi-elasticities (in order to estimate percentage changes in the dependent variable):

$$\log(y_i) = \log(\hat{y}_i) | x_i = (\hat{p}_i | x_i) \cdot (\log(\hat{y}_i) | y_i > 0, x_i) \quad (14)$$

Given a retransformed semi-elasticity from Equation 14, a change in the associated dummy variable from 0 to 1 corresponds to the following percentage change in the dependent variable²³⁴:

$$100 \cdot (e^\beta - 1) \quad (15)$$

Finally, note that all regressions in this Article, whether OLS, GLM, or two-part, are conducted using standard errors clustered by agency.

H. Choice of GLM Link Function and Distribution Family

Each GLM (whether standalone or second-step) requires explicit specification of the appropriate link function and distribution family, allowing it to accurately model non-linearity and heteroskedasticity.²³⁵ Although GLM regression requires additional analysis and is somewhat more computationally demanding, it is more accurate and will produce better results when OLS assumptions of normality and homoskedasticity do not hold.

In order to determine the appropriate distribution family for the GLM steps (and the GLM regression on preamble length), I use a modified Park test. The modified Park test is conducted using regression means (predicted values) and residuals from a two-part regression specified using a particular link function (here, a log link) and a hypothesized distribution family (here, a Poisson or Gamma distribution). The modified Park test evaluates the manner in which the square of the regression residuals (the variance) varies with the natural logarithm of the regression's predicted values (the mean). If a distribution is homoskedastic, as is assumed in OLS regression, then there should be no relationship between variance and mean. A modified Park test uses the following model, where \hat{y}_i is the predicted value from a regression (in the case of two-part regression, the second part) and r_i is the residual from the regression (that is, $r_i = y_i - \hat{y}_i$):

$$r_i^2 = \theta_i \cdot \log(\hat{y}_i) + \epsilon_i \quad (16)$$

The model in Equation 16 is evaluated in a GLM regression with parameters matching the initial hypothesized regression (a log link and either a Gamma or Poisson distribution family).

234. Eyal Frank, *Log-Linear Regressions: Three Things To Keep In Mind*, EYAL FRANK (Aug. 22, 2015), <http://www.eyalfrank.com/log-linear-regressions-three-things-to-keep-in-mind> [https://perma.cc/735R-XJP6].

235. However, standard GLM only allows us to model heteroskedasticity that itself follows one of several functional forms: Gaussian, Poisson, Gamma, or Wald (also known as inverse-Gaussian).

The appropriate distribution family for the GLM regression can be evaluated based on the estimated value of θ_i . If $\theta_i \approx 0$, that implies no relationship between mean and variance, so that a Gaussian distribution should be used, as with OLS regression. If $\theta_i \approx 1$, then a Poisson distribution should be used. If $\theta_i \approx 2$, then a Gamma distribution should be used. And if $\theta_i \approx 3$ then a Wald distribution (also known as an inverse-Gaussian distribution) should be used.²³⁶ Table 12 shows the results of the modified Park tests and the implied distribution families. The fact that each regression uses a Gamma distribution family affirms the choice not to use OLS in the second step of two-part regression.

Table 12: Results of Modified Park Tests and Distribution Family

Dependent Variable	θ_i	Distribution Family
Normative	2.308	Gamma
Statutory	2.504	Gamma
Legislative History	1.884	Gamma
Preamble Length	1.539	Gamma
Commenter	1.602	Gamma

Because the non-zero data in this study are log-normal,²³⁷ I use a log-link function (reflected by the log-transformations of the dependent variables in the equations above. In order to support that the log link was the appropriate choice, I plot residuals against the linear predictor for each GLM regression.²³⁸ If GLM assumptions are met, then the conditional mean function should be constant across the residual plot.²³⁹

236. Deb & Norton, *supra* note 226, at 497.

237. *See supra* Appendix Section F.

238. The trend lines are produced with locally estimated scatterplot smoothing.

239. JOHN FOX & SANFORD WEISBERG, AN R COMPANION TO APPLIED REGRESSION 320 (2011).

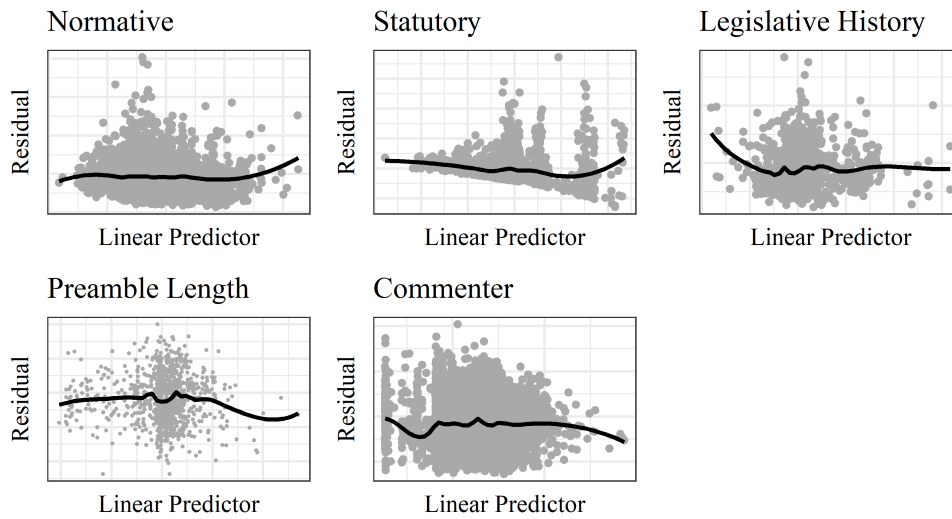
Figure 15: Residual Plots by GLM Dependent Variable

Figure 15 shows that the conditional means remain roughly constant across each residual plot, with deviations only at the extremes, driven by outliers. This supports the use of the log link in each of the GLM regressions and suggests that OLS regression (which assumes a Gaussian distribution family and an identity link) would have been inappropriate.

I. Regression Tables

This Section of the Appendix contains full results for the regressions described in this Article. Note that while many of the key interaction term coefficients are statistically significant and large in magnitude, the R^2 values (including McFadden's pseudo- R^2 , an analogous statistic for logistic and GLM models that reflects the improvement in the fitted model's performance over a naive baseline model) are relatively low. Intuitively, this means that although *Mayo* did have a substantial effect, the model is far from providing perfect predictions of preamble language. This makes sense: we should expect the actual substance of any regulation, rather than various summary statistics, to primarily drive the language chosen for the preamble.

For legibility, some of the tables are divided across two pages, as noted.

Table 13: OLS Regression Results for Normative and Statutory Terms, 2000-2020

Variables	(1) Normative	(2) Normative	(3) Normative	(4) Statutory	(5) Statutory	(6) Statutory
Post- <i>Mayo</i>	12.70 (26.24)	-32,450*** (9,007)	-38,241** (15,270)	28.03 (17.69)	8560 (7047)	5713 (7757)
Interpretative Tax Reg	-73.35* (37.39)	-69.85* (37.30)	-134.6*** (34.52)	47.28** (23.37)	45.97* (24.05)	22.41 (24.34)
Post- <i>Mayo</i> x Interpretative	104.8*** (26.24)	103.4*** (24.42)	104.2*** (26.20)	-77.47*** (17.69)	-76.38*** (17.32)	-82.74*** (15.83)
Year		-9.339** (4.120)	-8.882** (4.353)		3.510 (2.786)	3.598 (2.738)
Year x Post- <i>Mayo</i>		16.16*** (4.480)	19.02** (7.584)		-4.252 (3.507)	-2.841 (3.859)
Regulation Length			0.0299*** (0.00736)			0.00699** (0.00287)
High Reading Level			128.7*** (35.26)			-14.45 (25.18)
Democratic Administration			28.07 (23.59)			11.02 (9.489)
Constant	181.4*** (37.39)	18,904** (8,263)	17,826** (8,739)	91.34*** (23.37)	-6946 (5570)	-7073 (5481)
Fixed Effects	No	No	Yes	No	No	Yes
<i>N</i>	69,489	69,489	69,489	69,489	69,489	69,489
<i>R</i> ²	0.000	0.002	0.017	0.000	0.001	0.002

Note: Each column reports the results of a regression with the term frequency of either normative or statutory terms as the dependent variable, as indicated. Term frequency is measured in terms per million words. Standard errors are clustered by agency. *** p<0.01, ** p<0.05, * p<0.1

**Table 14: Two-Part Regression Results for Normative Terms, 2000-2020
(Part 1, Continued on Next Page)**

Variables	(1) Logit	(2) GLM	(3) Avg	(4) Logit	(5) GLM
Post- <i>Mayo</i>	0.219* (0.123)	-0.109** (0.0519)	0.0680 (0.109)	-43.26 (40.18)	-135.0*** (22.96)
Interpretative Tax Reg	-0.260 (0.305)	-0.300*** (0.0763)	-0.510** (0.246)	-0.257 (0.310)	-0.307*** (0.0689)
Post- <i>Mayo</i> x Interpretative	0.591*** (0.123)	0.199*** (0.0519)	0.676*** (0.126)	0.592*** (0.116)	0.181*** (0.0497)
Year				-0.00784 (0.0187)	-0.0433*** (0.0108)
Year x Post- <i>Mayo</i>				0.0216 (0.0200)	0.0672*** (0.0114)
Regulation Length					
High Reading Level					
Democratic Administration					
Constant	-1.530*** (0.305)	6.927*** (0.0763)		14.18 (37.39)	93.64*** (21.60)
Fixed Effects	No	No	No	No	No
<i>N</i>	69,489	69,489	69,489	69,489	69,489
McFadden's <i>R</i> ²	0.002	0.003		0.002	0.003

Note: The dependent variable is measured in terms per million words. The “Logit” and “GLM” columns reflect the first and second part of a two-part regression, while the “Combined” column reflects the semi-elasticity calculated by combining both columns. Standard errors are clustered by agency. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

**Table 14: Two-Part Regression Results for Normative Terms, 2000-2020
(Part 2, Continued from Previous Page)**

Variables	(6) Avg	(7) Logit	(8) GLM	(9) Avg
Post- <i>Mayo</i>	-169.9*** (40.59)	-51.12 (78.74)	-142.2*** (33.29)	-183.5** (72.93)
Interpretative Tax Reg	-0.514** (0.247)	-0.999*** (0.219)	-0.205*** (0.0504)	-1.011*** (0.159)
Post- <i>Mayo</i> x Interpretative	0.659*** (0.119)	0.819*** (0.215)	0.204*** (0.0412)	0.864*** (0.212)
Year	-0.0496*** (0.0188)	-0.00626 (0.0225)	-0.0406*** (0.0102)	-0.0456** (0.0210)
Year x Post- <i>Mayo</i>	0.0846*** (0.0202)	0.0255 (0.0391)	0.0707*** (0.0165)	0.0913** (0.0362)
Regulation Length		0.000210*** (6.40e-05)	-2.35e-05** (1.05e-05)	0.000146** (6.22e-05)
High Reading Level		2.178*** (0.439)	-0.664*** (0.150)	1.092*** (0.287)
Democratic Administration		0.0753 (0.0968)	0.0612 (0.0496)	0.122 (0.0947)
Constant		8.249 (45.31)	89.18*** (20.39)	
Fixed Effects	No	Yes	Yes	Yes
<i>N</i>	69,489	69,489	69,489	69,489
McFadden's <i>R</i> ²		0.080	0.003	

Note: The dependent variable is measured in terms per million words. The “Logit” and “GLM” columns reflect the first and second part of a two-part regression, while the “Combined” column reflects the semi-elasticity calculated by combining both columns. Standard errors are clustered by agency. *** p<0.01, ** p<0.05, * p<0.1

Table 15: Two-Part Regression Results for Statutory Terms, 2000-2020 (Part 1, Continued on Next Page)

Variables	(1) Logit	(2) GLM	(3) Avg	(4) Logit	(5) GLM
Post- <i>Mayo</i>	0.437*** (0.147)	-0.128** (0.0512)	0.268** (0.134)	7.478 (54.72)	79.92* (42.98)
Interpretative Tax Reg	0.747** (0.300)	-0.246* (0.145)	0.429 (0.327)	0.746** (0.297)	-0.262* (0.140)
Post- <i>Mayo</i> x Interpretative	-0.467*** (0.147)	-0.288*** (0.0512)	-0.710*** (0.134)	-0.466*** (0.149)	-0.272*** (0.0555)
Year				0.00309 (0.0211)	0.0345** (0.0143)
Year x Post- <i>Mayo</i>				-0.00351 (0.0273)	-0.0399* (0.0214)
Regulation Length					
High Reading Level					
Democratic Administration					
Constant	-2.466*** (0.300)	7.062*** (0.145)		-8.657 (42.35)	-62.05** (28.71)
Fixed Effects	No	No	No	No	No
<i>N</i>	69,489	69,489	69,489	69,489	69,489
McFadden's <i>R</i> ²	0.007	0.010		0.007	0.010

Note: The dependent variable is measured in terms per million words. The “Logit” and “GLM” columns reflect the first and second part of a two-part regression, while the “Combined” column reflects the semi-elasticity calculated by combining both columns. Standard errors are clustered by agency. *** p<0.01, ** p<0.05, * p<0.1

Table 15: Two-Part Regression Results for Statutory Terms, 2000-2020 (Part 2, Continued from Previous Page)

Variables	(6) Avg	(7) Logit	(8) GLM	(9) Avg
Post- <i>Mayo</i>	86.68 (65.60)	-30.99 (82.05)	92.16 (57.65)	64.15 (93.47)
Interpretative Tax Reg	0.412 (0.322)	0.287 (0.184)	0.143 (0.122)	0.402* (0.212)
Post- <i>Mayo</i> x Interpretative	-0.693*** (0.137)	-0.474*** (0.137)	-0.235*** (0.0423)	-0.664*** (0.122)
Year	0.0373 (0.0239)	0.00648 (0.0210)	0.0254 (0.0195)	0.0313 (0.0273)
Year x Post- <i>Mayo</i>	-0.0431 (0.0327)	0.0155 (0.0408)	-0.0459 (0.0287)	-0.0318 (0.0465)
Regulation Length		0.000153*** (3.75e-05)	-7.71e-05*** (1.44e-05)	6.14e-05 (4.06e-05)
High Reading Level		1.584*** (0.274)	-1.578*** (0.225)	-0.147 (0.316)
Democratic Administration		0.194* (0.0997)	-0.0773** (0.0385)	0.0984 (0.0950)
Constant		-17.41 (41.98)	-41.90 (39.28)	
Fixed Effects	No	Yes	Yes	Yes
<i>N</i>	69,489	69,489	69,489	69,489
McFadden's <i>R</i> ²		0.050	0.006	

Note: The dependent variable is measured in terms per million words. The “Logit” and “GLM” columns reflect the first and second part of a two-part regression, while the “Combined” column reflects the semi-elasticity calculated by combining both columns. Standard errors are clustered by agency. *** p<0.01, ** p<0.05, * p<0.1

Table 16: OLS Regression Results for Preamble Length and Commenter Terms, 2000-2020

Variables	(1) Preamble	(2) Preamble	(3) Preamble	(4) Comm- enter	(5) Comm- enter	(6) Comm- enter
Post- <i>Mayo</i>	172.1* (93.12)	20,481 (21,485)	7509 (23,715)	-139.4 (167.4)	-70,460 (68,175)	-72,120 (56,963)
Interpretative Tax Reg	159.5 (172.2)	152.7 (171.7)	-486.4*** (153.0)	931.1*** (203.9)	941.2*** (193.0)	642.4*** (218.0)
Post- <i>Mayo</i> x Interpretative	411.7*** (93.12)	420.8*** (88.58)	439.7*** (67.73)	202.1 (167.4)	194.4 (156.4)	216.6 (169.3)
Year		17.83** (6.995)	19.51** (7.844)		-26.96 (32.51)	-28.09 (32.58)
Year x Post- <i>Mayo</i>		-10.17 (10.69)	-3.766 (11.79)		35.03 (33.91)	35.86 (28.35)
Regulation Length			0.284*** (0.0285)			0.147*** (0.0249)
High Reading Level			903.5*** (162.4)			316.6*** (84.75)
Democratic Administration			92.63*** (28.67)			1.484 (40.87)
Constant	1273*** (172.2)	-34,476** (14,062)	-38,831** (15,750)	646.7*** (203.9)	54,701 (65,365)	56,306 (65,384)
Fixed Effects	No	No	Yes	No	No	Yes
<i>N</i>	69,575	69,575	69,575	69,489	69,489	69,489
<i>R</i> ²	0.003	0.003	0.106	0.006	0.007	0.033

Note: Each column reports the results of a regression with the preamble word count or term frequency of references to commenters as the dependent variable, as indicated. Term frequency is measured in terms per million words. Standard errors are clustered by agency. *** p<0.01, ** p<0.05, * p<0.1

Table 17: GLM Regression Results for Preamble Length, 2000-2020

Variables	(1)	(2)	(3)	(4)
Post- <i>Mayo</i>	0.127* (0.0739)	18.42 (13.59)	16.88 (17.78)	16.91 (17.78)
Interpretative Tax Reg	0.118 (0.135)	0.118 (0.136)	-0.580*** (0.101)	15.55 (13.47)
Post- <i>Mayo</i> x Interpretative	0.215*** (0.0739)	0.215*** (0.0746)	0.173** (0.0702)	0.259*** (0.0283)
Year		0.0140** (0.00618)	0.0161** (0.00757)	0.0134** (0.00544)
Year x Post- <i>Mayo</i>		-0.00915 (0.00676)	-0.00840 (0.00884)	-0.00842 (0.00884)
Regulation Length			0.000191*** (2.77e-05)	0.000191*** (2.78e-05)
High Reading Level			1.178*** (0.185)	1.178*** (0.185)
Democratic Administration			0.0645*** (0.0196)	0.0644*** (0.0196)
Year x Pre-Trend Slope				0.000146 (0.000121)
Constant	7.149*** (0.135)	-20.83* (12.45)	-26.49* (15.20)	-26.72* (15.36)
Fixed Effects	No	No	Yes	Yes
<i>N</i>	69,575	69,575	69,575	69,575
McFadden's <i>R</i> ²	0.009	0.009	0.007	0.007

Note: The dependent variable in each column is the number of words per regulatory preamble. Standard errors are clustered by agency. *** p<0.01, ** p<0.05, * p<0.1

**Table 18: Two-Part Regression Results for Commenter Terms, 2000-2020
(Part 1, Continued on Next Page)**

Variables	(1) Logit	(2) GLM	(3) Avg	(4) Logit	(5) GLM
Post- <i>Mayo</i>	-0.0359 (0.179)	-0.213* (0.119)	-0.243 (0.188)	-94.10 (72.31)	-40.05 (36.03)
Interpretative Tax Reg	1.073*** (0.262)	0.124 (0.117)	0.997*** (0.271)	1.084*** (0.248)	0.129 (0.108)
Post- <i>Mayo</i> x Interpretative	-0.0307 (0.179)	0.293** (0.119)	0.268 (0.189)	-0.0339 (0.166)	0.288*** (0.108)
Year				-0.0248 (0.0337)	-0.0224 (0.0156)
Year x Post- <i>Mayo</i>				0.0468 (0.0360)	0.0199 (0.0179)
Regulation Length					
High Reading Level					
Democratic Administration					
Constant	-1.481*** (0.262)	8.158*** (0.117)		48.15 (67.76)	53.12* (31.30)
Fixed Effects	No	No	No	No	No
<i>N</i>	69,489	69,489	69,489	69,489	69,489
McFadden's R^2	0.004	0.005		0.005	0.005

Note: The dependent variable is measured in terms per million words. The “Logit” and “GLM” columns reflect the first and second part of a two-part regression, while the “Combined” column reflects the semi-elasticity calculated by combining both columns. Standard errors are clustered by agency. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

**Table 18: Two-Part Regression Results for Commenter Terms, 2000-2020
(Part 2, Continued from Previous Page)**

Variables	(6) Avg	(7) Logit	(8) GLM	(9) Avg
Post- <i>Mayo</i>	-116.6* (67.52)	-85.06 (62.25)	-47.85* (28.80)	-117.1** (57.12)
Interpretative Tax Reg	1.010*** (0.258)	0.728*** (0.280)	0.0532 (0.116)	0.646** (0.272)
Post- <i>Mayo</i> x Interpretative	0.260 (0.173)	0.0109 (0.177)	0.298*** (0.111)	0.306* (0.182)
Year	-0.0426 (0.0311)	-0.0245 (0.0347)	-0.0228 (0.0165)	-0.0428 (0.0323)
Year x Post- <i>Mayo</i>	0.0580* (0.0336)	0.0423 (0.0310)	0.0238* (0.0143)	0.0582** (0.0284)
Regulation Length		0.000184*** (3.10e-05)	2.52e-05** (1.18e-05)	0.000175*** (2.77e-05)
High Reading Level		1.274*** (0.341)	-0.258*** (0.0889)	0.778** (0.315)
Democratic Administration		-0.00995 (0.0570)	0.00466 (0.0652)	-0.00343 (0.0799)
Constant		45.86 (70.10)	53.82 (32.95)	
Fixed Effects	No	Yes	Yes	Yes
<i>N</i>	69,489	69,489	69,489	69,489
McFadden's <i>R</i> ²		0.060	0.005	

Note: The dependent variable is measured in terms per million words. The “Logit” and “GLM” columns reflect the first and second part of a two-part regression, while the “Combined” column reflects the semi-elasticity calculated by combining both columns. Standard errors are clustered by agency. *** p<0.01, ** p<0.05, * p<0.1

Table 19: Two-Part Regression Results for Legislative History Terms, 2000-2020 (Part 1, Continued on Next Page)

Variables	(1) Logit	(2) GLM	(3) Avg	(4) Logit	(5) GLM
Post- <i>Mayo</i>	-0.263** (0.104)	-0.169** (0.0860)	-0.429*** (0.134)	0.0408 (63.01)	30.60 (69.13)
Interpretative Tax Reg	2.382*** (0.276)	0.435*** (0.0858)	2.783*** (0.292)	2.387*** (0.275)	0.434*** (0.0844)
Post- <i>Mayo</i> x Interpretative	-0.222** (0.104)	-0.0510 (0.0860)	-0.270** (0.134)	-0.230** (0.105)	-0.000241 (0.0732)
Year				-0.0121 (0.0185)	-0.0290* (0.0172)
Year x Post- <i>Mayo</i>				-8.74e-05 (0.0313)	-0.0151 (0.0343)
Regulation Length					
High Reading Level					
Democratic Administration					
Constant	-4.247*** (0.276)	6.602*** (0.0858)		20.09 (37.16)	64.76* (34.51)
Fixed Effects	No	No	No	No	No
<i>N</i>	69,489	69,489	69,489	69,489	69,489
McFadden's <i>R</i> ²	0.034	0.007		0.034	0.007

Note: The dependent variable is measured in terms per million words. The “Logit” and “GLM” columns reflect the first and second part of a two-part regression, while the “Combined” column reflects the semi-elasticity calculated by combining both columns. Standard errors are clustered by agency. *** p<0.01, ** p<0.05, * p<0.1

Table 19: Two-Part Regression Results for Legislative History Terms, 2000-2020 (Part 2, Continued from Previous Page)

Variables	(6) Avg	(7) Logit	(8) GLM	(9) Avg
Post- <i>Mayo</i>	30.64 (92.93)	-15.70 (62.74)	3.150 (52.18)	-12.33 (80.90)
Interpretative Tax Reg	2.787*** (0.291)	1.959*** (0.230)	0.656*** (0.0712)	2.587*** (0.243)
Post- <i>Mayo</i> x Interpretative	-0.227* (0.127)	-0.261*** (0.0926)	0.0238 (0.0862)	-0.234* (0.126)
Year	-0.0410 (0.0251)	-0.00399 (0.0153)	-0.0200 (0.0130)	-0.0239 (0.0199)
Year x Post- <i>Mayo</i>	-0.0152 (0.0462)	0.00764 (0.0312)	-0.00155 (0.0259)	0.00598 (0.0402)
Regulation Length		0.000146*** (2.39e-05)	-5.51e-05*** (1.04e-05)	8.93e-05*** (2.59e-05)
High Reading Level		2.446*** (0.334)	-1.220*** (0.463)	1.191** (0.569)
Democratic Administration		0.169* (0.101)	0.0388 (0.111)	0.205 (0.149)
Constant		0.916 (30.71)	48.25* (26.13)	
Fixed Effects	No	Yes	Yes	Yes
<i>N</i>	69,489	69,489	69,489	69,489
McFadden's <i>R</i> ²		0.073	0.005	

Note: The dependent variable is measured in terms per million words. The “Logit” and “GLM” columns reflect the first and second part of a two-part regression, while the “Combined” column reflects the semi-elasticity calculated by combining both columns. Standard errors are clustered by agency. *** p<0.01, ** p<0.05, * p<0.1

Table 20: Shortened-Window Regression Results, 2006-2015 (Part 1, Continued on Next Page)

Variables	(1) Normative Logit	(2) Normative GLM	(3) Normative Avg	(4) Statutory Logit	(5) Statutory GLM
Post- <i>Mayo</i>	33.05 (86.18)	-73.85 (60.03)	-47.09 (91.79)	27.10 (166.9)	93.22 (100.9)
Interpretative Tax Reg	-0.949*** (0.261)	-0.235*** (0.0615)	-1.004*** (0.198)	0.375** (0.180)	0.171 (0.170)
Post- <i>Mayo</i> x Interpretative	0.442** (0.213)	0.134** (0.0525)	0.493** (0.197)	-0.947*** (0.202)	-0.242*** (0.0620)
Year	-0.00276 (0.0376)	-0.0237 (0.0285)	-0.0260 (0.0417)	0.0369 (0.0323)	0.0306 (0.0442)
Year x Post- <i>Mayo</i>	-0.0164 (0.0429)	0.0367 (0.0299)	0.0235 (0.0457)	-0.0133 (0.0830)	-0.0464 (0.0502)
Regulation Length	0.000209*** (6.81e-05)	-2.29e-05* (1.29e-05)	0.000146** (6.61e-05)	0.000167*** (3.62e-05)	-7.43e-05*** (1.63e-05)
High Reading Level	2.215*** (0.418)	-0.557*** (0.156)	1.236*** (0.275)	1.746*** (0.250)	-1.711*** (0.270)
Democratic Administration	0.177** (0.0804)	-0.0481 (0.0913)	0.0951 (0.109)	-0.0190 (0.148)	-0.104 (0.0882)
Constant	1.212 (75.73)	55.15 (57.18)		-78.33 (64.92)	-51.93 (88.69)
Fixed Effects	Yes	Yes	Yes	Yes	Yes
<i>N</i>	32,424	32,424	32,424	32,424	32,424
McFadden's R^2	0.086	0.002		0.057	0.006

Note: The dependent variable is measured in terms per million words (except for preamble length, which measures word count). The “Logit” and “GLM” columns reflect the first and second part of a two-part regression, while the “Combined” column reflects the semi-elasticity calculated by combining both columns. Standard errors are clustered by agency. *** p<0.01, ** p<0.05, * p<0

Table 20: Shortened-Window Regression Results, 2006-2015 (Part 2, Continued from Previous Page)

Variables	(6) Statutory Avg	(7) Preamble GLM	(8) Commenter Logit	(9) Commenter GLM	(10) Commenter Avg
Post- <i>Mayo</i>	117.6 (181.3)	30.76 (28.90)	-136.9 (118.5)	1.388 (49.93)	-111.4 (108.2)
Interpretative Tax Reg	0.509** (0.240)	-0.661*** (0.0949)	0.764*** (0.194)	0.130* (0.0707)	0.759*** (0.187)
Post- <i>Mayo</i> x Interpretative	-1.094*** (0.175)	0.113*** (0.0371)	-0.0349 (0.101)	0.188*** (0.0510)	0.159 (0.0981)
Year	0.0637 (0.0531)	0.0139 (0.0129)	-0.0294 (0.0407)	-0.0194 (0.0246)	-0.0437 (0.0413)
Year x Post- <i>Mayo</i>	-0.0584 (0.0902)	-0.0153 (0.0144)	0.0681 (0.0589)	-0.000730 (0.0248)	0.0554 (0.0538)
Regulation Length	7.56e-05* (4.03e-05)	0.000188*** (2.88e-05)	0.000177*** (3.34e-05)	1.94e-05 (1.34e-05)	0.000165*** (3.08e-05)
High Reading Level	-0.140 (0.344)	1.156*** (0.199)	1.073** (0.452)	-0.175* (0.105)	0.709* (0.398)
Democratic Administration	-0.121 (0.160)	0.0373 (0.0403)	-0.0291 (0.0763)	0.130 (0.0808)	0.106 (0.103)
Constant		-21.97 (25.98)	55.99 (82.18)	46.97 (49.26)	
Fixed Effects	Yes	Yes	Yes	Yes	Yes
<i>N</i>	32,424	32,439	32,424	32,424	32,424
McFadden's <i>R</i> ²		0.007	0.056	0.005	

Note: The dependent variable is measured in terms per million words (except for preamble length, which measures word count). The “Logit” and “GLM” columns reflect the first and second part of a two-part regression, while the “Combined” column reflects the semi-elasticity calculated by combining both columns. Standard errors are clustered by agency. *** p<0.01, ** p<0.05, * p<0

Table 21: Winsorized Regression Results, 2000-2020 (Part 1, Continued on Next Page)

Variables	(1) Normative Logit	(2) Normative GLM	(3) Normative Avg	(4) Statutory Logit	(5) Statutory GLM
Post- <i>Mayo</i>	-51.12 (78.74)	-88.93*** (22.53)	-130.2* (68.68)	-30.99 (82.05)	21.21 (18.32)
Interpretative Tax Reg	-0.999*** (0.219)	-0.138*** (0.0306)	-0.944*** (0.154)	0.287 (0.184)	0.115** (0.0491)
Post- <i>Mayo</i> x Interpretative	0.819*** (0.215)	0.125*** (0.0329)	0.785*** (0.211)	-0.474*** (0.137)	-0.144*** (0.0190)
Year	-0.00626 (0.0225)	-0.0285*** (0.00630)	-0.0336* (0.0194)	0.00648 (0.0210)	0.00119 (0.00636)
Year x Post- <i>Mayo</i>	0.0255 (0.0391)	0.0442*** (0.0112)	0.0648* (0.0341)	0.0155 (0.0408)	-0.0105 (0.00912)
Regulation Length	0.000210*** (6.40e-05)	-2.13e-05*** (4.87e-06)	0.000148** (6.15e-05)	0.000153*** (3.75e-05)	-4.34e-05*** (5.30e-06)
High Reading Level	2.178*** (0.439)	-0.224*** (0.0627)	1.533*** (0.252)	1.584*** (0.274)	-0.320*** (0.0612)
Democratic Administration	0.0753 (0.0968)	0.00694 (0.0377)	0.0677 (0.0890)	0.194* (0.0997)	-0.0476*** (0.0185)
Constant	8.249 (45.31)	64.26*** (12.65)		-17.41 (41.98)	4.502 (12.79)
Fixed Effects	Yes	Yes	Yes	Yes	Yes
<i>N</i>	69,489	69,489	69,489	69,489	69,489
McFadden's <i>R</i> ²	0.080	0.001		0.050	0.001

Note: The dependent variable is measured in terms per million words (except for preamble length, which measures word count). The “Logit” and “GLM” columns reflect the first and second part of a two-part regression, while the “Combined” column reflects the semi-elasticity calculated by combining both columns. Standard errors are clustered by agency. *** p<0.01, ** p<0.05, * p<0.1

Table 21: Winsorized Regression Results, 2000-2020 (Part 2, Continued from Previous Page)

Variables	(6) Statutory Avg	(7) Preamble GLM	(8) Commenter Logit	(9) Commenter GLM	(10) Commenter Avg
Post- <i>Mayo</i>	-6.802 (75.82)	21.65 (17.50)	-85.06 (62.25)	-22.72 (21.75)	-91.94* (53.90)
Interpretative Tax Reg	0.374** (0.180)	-0.494*** (0.0946)	0.728*** (0.280)	0.171* (0.0885)	0.763*** (0.262)
Post- <i>Mayo</i> x Interpretative	-0.573*** (0.116)	0.140** (0.0635)	0.0109 (0.177)	0.212*** (0.0683)	0.221 (0.159)
Year	0.00704 (0.0201)	0.0142* (0.00767)	-0.0245 (0.0347)	-0.0102 (0.00920)	-0.0302 (0.0292)
Year x Post- <i>Mayo</i>	0.00351 (0.0377)	-0.0108 (0.00870)	0.0423 (0.0310)	0.0113 (0.0108)	0.0457* (0.0268)
Regulation Length	9.51e-05** (3.83e-05)	0.000171*** (2.75e-05)	0.000184*** (3.10e-05)	1.83e-05* (9.86e-06)	0.000168*** (2.69e-05)
High Reading Level	1.111*** (0.230)	1.169*** (0.174)	1.274*** (0.341)	-0.227*** (0.0557)	0.809*** (0.308)
Democratic Administration	0.128 (0.0888)	0.0490*** (0.0166)	-0.00995 (0.0570)	-0.00998 (0.0485)	-0.0181 (0.0670)
Constant		-22.64 (15.37)	45.86 (70.10)	28.37 (18.34)	
Fixed Effects	Yes	Yes	Yes	Yes	Yes
<i>N</i>	69,489	69,575	69,489	69,489	69,489
McFadden's <i>R</i> ²		0.005	0.060	0.003	

Note: The dependent variable is measured in terms per million words (except for preamble length, which measures word count). The “Logit” and “GLM” columns reflect the first and second part of a two-part regression, while the “Combined” column reflects the semi-elasticity calculated by combining both columns. Standard errors are clustered by agency. *** p<0.01, ** p<0.05, * p<0.1

Table 22: Placebo Regression Results, 2000-2010 (Part 1, Continued on Next Page)

Variables	(1) Normative Logit	(2) Normative GLM	(3) Normative Avg	(4) Statutory Logit	(5) Statutory GLM
Post-2005	-41.07 (85.56)	-58.80 (63.48)	-92.60 (94.78)	-137.2 (84.03)	30.40 (83.17)
Interpretative Tax Reg	-1.045*** (0.213)	-0.155** (0.0645)	-1.015*** (0.160)	0.103 (0.218)	0.169** (0.0844)
Post- <i>Mayo</i> x Interpretative	0.137 (0.0834)	-0.00545 (0.0663)	0.107 (0.0980)	0.235 (0.144)	0.0269 (0.121)
Year	-0.00413 (0.0302)	-0.0654*** (0.0181)	-0.0688** (0.0308)	-0.0354 (0.0233)	0.0432** (0.0186)
Year x Post- 2005	0.0204 (0.0426)	0.0294 (0.0316)	0.0462 (0.0472)	0.0685 (0.0419)	-0.0152 (0.0414)
Regulation Length	0.000213*** (5.77e-05)	-3.76e-05*** (1.16e-05)	0.000138** (5.61e-05)	0.000170*** (3.55e-05)	-8.66e-05*** (1.30e-05)
High Reading Level	2.024*** (0.432)	-0.792*** (0.164)	0.873*** (0.323)	1.298*** (0.288)	-1.503*** (0.263)
Democratic Administration	0.112 (0.132)	0.0896 (0.0667)	0.182 (0.128)	0.0599 (0.149)	-0.0464 (0.0699)
Constant	3.498 (60.62)	138.9*** (36.28)		66.74 (46.47)	-77.59** (37.38)
Fixed Effects	Yes	Yes	Yes	Yes	Yes
<i>N</i>	36,514	36,514	36,514	36,514	36,514
McFadden's <i>R</i> ²	0.076	0.003		0.040	0.006

Note: The dependent variable is measured in terms per million words (except for preamble length, which measures word count). The “Logit” and “GLM” columns reflect the first and second part of a two-part regression, while the “Combined” column reflects the semi-elasticity calculated by combining both columns. Standard errors are clustered by agency. *** p<0.01, ** p<0.05, * p<0.1

Table 22: Placebo Regression Results, 2000-2010 (Part 2, Continued from Previous Page)

Variables	(6) Statutory Avg	(7) Preamble GLM	(8) Commenter Logit	(9) Commenter GLM	(10) Commenter Avg
Post-2005	-95.89 (114.4)	-8.411 (54.65)	130.0* (66.54)	171.2* (97.33)	276.5** (112.2)
Interpretative Tax Reg	0.264 (0.220)	-0.534*** (0.102)	0.644* (0.335)	0.0180 (0.175)	0.539 (0.344)
Post- <i>Mayo</i> x Interpretative	0.244 (0.178)	-0.162*** (0.0615)	0.110 (0.157)	0.138 (0.113)	0.227 (0.168)
Year	0.0106 (0.0285)	0.0119 (0.0152)	0.0240 (0.0378)	0.0571*** (0.0200)	0.0765*** (0.0372)
Year x Post- 2005	0.0478 (0.0570)	0.00420 (0.0273)	-0.0649* (0.0332)	-0.0856* (0.0486)	-0.138** (0.0560)
Regulation Length	6.99e-05* (3.80e-05)	0.000206*** (3.93e-05)	0.000192*** (3.31e-05)	1.97e-05 (1.59e-05)	0.000176*** (3.18e-05)
High Reading Level	-0.308 (0.367)	1.320*** (0.217)	1.397*** (0.254)	-0.352*** (0.0969)	0.779*** (0.265)
Democratic Administration	0.00878 (0.153)	0.0213 (0.0410)	0.0491 (0.0596)	0.154*** (0.0554)	0.194*** (0.0742)
Constant		-18.24 (30.50)	-51.53 (76.33)	-106.4*** (40.24)	
Fixed Effects	Yes	Yes	Yes	Yes	Yes
<i>N</i>	36,514	36,597	36,514	36,514	36,514
McFadden's <i>R</i> ²		0.007	0.062	0.005	

Note: The dependent variable is measured in terms per million words (except for preamble length, which measures word count). The “Logit” and “GLM” columns reflect the first and second part of a two-part regression, while the “Combined” column reflects the semi-elasticity calculated by combining both columns. Standard errors are clustered by agency. *** p<0.01, ** p<0.05, * p<0.1