EFFICACY OF CONGRESSIONAL OVERSIGHT

Pamela Ban* Seth J. Hill†

August 22, 2023

Abstract: Scholars argue that oversight allows Congress to control the executive agents it empowers to implement law. Yet the tools of oversight are rather limited and debate continues as to how much political control oversight provides. How well can members of Congress motivate action within the bureaucracy? To measure the efficacy of oversight, we create a new data set on a bureaucratic deficiency that Congress has sought to reduce since the early 2000s: improperly-made payments to contractors and clients. We estimate the effect of congressional hearings, one of the most important tools of congressional oversight, on subsequent improper payments. We find that hearings on the issue do lead to a decline in improper payments for agencies whose employees are called to testify. But the magnitude of the effect is small relative to the base rate, suggesting strong limits on the effectiveness of congressional oversight. We find similarly small or no effects of correspondence, appropriation committee reports, statutes, and executive action. Our findings strongly imply that America's elected officials struggle to effectively manage implementation of government policy.

Keywords: Congressional oversight; political control; principal-agent problem; bureaucratic politics.

*University of California San Diego, 9500 Gilman Drive #0521, La Jolla, CA 92093-0521; pmban@ucsd.edu, www.pamelaban.com.

†University of California San Diego, 9500 Gilman Drive #0521, La Jolla, CA 92093-0521; sjhill@ucsd.edu, www.sethjhill.com.

> **It is easy to see how the commands as well as the questions of Congress may be evaded, if not directly disobeyed, by the executive agents. Its Committees may command, but they cannot superintend the execution of their commands.**
>
> (Wilson, 1885 [2018], p. 127)

At the heart of the production of American public policy lies a principal-agent problem. Congress passes laws that empower executive agencies to produce goods and services and to direct the behavior of citizens and firms. The ability of the legislative principal to monitor and control the efforts of the bureaucratic agent, however, is circumscribed by informational and contractual challenges. The congressional principal has a hard time learning about and evaluating the agent's choice of action. And even when legislators learn that agents act contrary to their preferences, civil service protections and separation of powers make it difficult for the legislature to direct the actions taken by employees of the executive branch.

A long literature in political science explores the many ways that legislatures work within these challenges. Debate has focused on whether the legislature is on balance successful in controlling bureaucratic behavior (e.g., Weingast and Moran, 1983; McCubbins, Noll, and Weingast, 1987; McCubbins and Schwartz, 1984) or whether the legislature's influence is weak (e.g., Moe, 1987). This debate remains unresolved after decades of research.

Among the tools used to limit agency loss, scholars most commonly discuss oversight. Oversight hearings, in particular, have been recognized as a process through members of Congress collect information on bureaucratic efforts, highlight malfeasance, communicate their preferences to members of the bureaucracy, and use the implied threat of future budgetary or statutory restrictions to motivate bureaucratic responsiveness to congressional wants (McCubbins and Schwartz, 1984; Aberbach, 1990; Kriner and Schwartz, 2008; McGrath, 2013; MacDonald and McGrath, 2016; Lowande and Peck, 2017). Previous work tells us about when Congress is likely to perform oversight (e.g., McGrath, 2013) and how the distribution of oversight authority across committees affects perceptions of Congress's ability to drive agency policy (e.g., Clinton, Lewis, and Selin, 2014).

Almost all empirical work measures the *presence* of oversight action, when and how Congress conducts oversight, but stops at the question of how effective oversight is on influencing agency behavior. Yet our theories of political control assume that oversight influences bureaucratic choices, making this an important bridge between theory and evidence to date unbuilt. How effective is congressional oversight?

We argue that this hole in our understanding is due to the inherent difficulty in evaluating the perfor-

mance of the federal bureaucracy. Because the federal bureaucracy often produces public goods that, by definition, are difficult to price through market mechanisms, it is hard to evaluate when and by how much bureaucratic efforts match legislative wants. For example, is a cost of $78 million per F-35 strike aircraft (Government Accountability Office, 2022) evidence of effective or ineffective oversight of the defense bureaucracy?

To estimate the effectiveness of congressional oversight, one would want measures of the goals of Congress and the output of the bureaucracy on a common scale, ideally measurable and comparable across agencies and over time. Comparison of the goals of Congress and the output of the bureaucracy on such a scale would provide evidence on the magnitude of control versus agency loss. Because agencies implement a range of programs that tackle diverse sets of issues with varied mandates, constraints, and institutional settings, a comparable measure is no easy task. Indeed, scholars of the bureaucracy have suggested that a comparable, quantitative measure of agency and program outcomes is one of the chief obstacles in the study of bureaucratic politics.[1] One approach has been to use questions on program performance asked in surveys of federal bureaucrats (e.g., Lewis, 2007; Gilmour and Lewis, 2006; Clinton, Lewis, and Selin, 2014), but these measures are limited in availability across time and might differ in application or measurement across programs and respondents.

Here we create a new data set of a plausibly-comparable measure of bureaucratic performance to estimate the efficacy of congressional oversight. Since the early 2000s, Congress has sought to reduce payments made by bureaucracies to contractors and clients deemed, subsequent to disbursement, improper. An improper payment is "any payment that should not have been made or that was made in an incorrect amount under statutory, contractual, administrative, or other legally applicable requirement" (Paymentaccuracy.gov, 2023). Examples of improper payments include Medicare payments for beneficiary goods (e.g., powered wheelchairs) without adequate prior authorization or for services determined in a subsequent audit not medically necessary, Medicaid payments to providers ineligible due to suspended or revoked licenses, Internal Revenue Service tax refunds paid without following required procedures to review W-2 and other documentation, Department of Homeland Security bills paid to a contractor for services beyond those authorized by statute, or Department of Labor unemployment benefits subsequently determined to be fraudulent. According to the Government Accountability Office (2023, GAO, p. 15), the main causes of improper payments are

[1] Lewis (2007, p. 1075) writes, "One major difficulty is that it is hard to define good performance objectively and in a manner acceptable to different stakeholders."

failure to access the information to properly document and disburse the payment, inability to access or non-existence of information needed to follow statute, or lack of documentation from applicants to determine payment eligibility. These problems mix employee behavior, accounting systems, information technology, and contractor or applicant actions (including fraud).

Improper payments can be made by any agency or program in any fiscal year. Because both improper payments and agency expenditures are measured in dollars, the fraction of all payments improper (the "improper payment rate") provides a plausibly comparable over-time and cross-agency measure of agency performance. Further, improper payments, totaling $247 billion in fiscal year 2022 and more than $2.4 trillion since 2003, are a large bureaucratic deficiency that have been a recurring target of oversight, executive directives, and statutes from members of both political parties.

Unlike many other outputs of the bureaucracy, for this outcome there is near-universal agreement on what Congress desires from the bureaucracy (fewer improper payments). This allows us to set aside bureaucratic policies where members of the legislature have different priorities, which complicates evaluation of political control, and instead focus on a valence issue with objective measures of what would (and would not) follow from effective legislative control. Congress wants fewer improper payments and, so, greater legislative control implies lower rates of improper payments.

We compile new data on the fiscal year rates of improper payments at the agency and program level and compare changes in this measure to the incidence of oversight. We examine the effect of congressional hearings, a keystone of legislative oversight. If oversight is effective in mitigating agency loss, we should observe that hearings lead to a decline in improper payment rates. If, on the other hand, the power of the legislative branch to motivate action in the executive is weak, we would not see such a decline. Congress has used oversight hearings in attempts to reduce levels and rates of improper payments for more than two decades, providing a unique opportunity to study the influence of congressional oversight on executive branch effort and output.

We imagine oversight as an informational exchange between the legislature and the executive. The bureaucracy has more potential activities consistent with statute than resources to pursue them and oversight communication allows members of Congress to express their priorities for bureaucratic choice of pursuit. Our read of hearing transcripts about improper payments is consistent with this interpretation. Members query officials about the actions they are taking and plan to take to combat improper payments and about what resources might help them be more effective. Officials learn from the statements and questions of

members what the members prioritize. Our research question, then, is what consequences these interbranch exchanges have on bureaucratic agency loss.

Should we expect oversight to motivate action within the bureaucracy? On one hand, information revealed through oversight activities can be used when Congress decides on program authorizations, budgets, or future policy proposals, possibly inducing agencies to respond to oversight. Scholarship suggests that other forms of nonstatutory control influence bureaucratic behavior (Bolton, 2022; Acs, 2019); oversight activities might similarly spur bureaucratic response. On the other hand, simply revealing information about an agency's performance does not, in isolation, require agencies to take action. The efficacy of oversight stems from the public revelation of agency performance and the implicit threat that Congress, in gaining this information, is watching the agency and might take future action. This limitation to oversight may thus narrow its effectiveness in influencing agency behavior (as argued by, e.g., Wilson, 1885 [2018]; Moe, 1987).

Oversight could even appear to have consequences opposite the desired effect. Change in agency performance may first require additional bureaucratic attention to the problem and time to implement reforms. If efficacy of oversight stems from the attention it brings to the problem, hearings held on improper payments may lead to a near-term increase in rates as the targeted agency increases staff time devoted to improper payments and discovers previously unappreciated errors in its payment portfolio.

One immediate concern with our design is that witnesses at oversight hearings are not invited at random. Instead, congressional committees might be selecting agencies or programs for oversight because of issues with payment integrity. Although we cannot rule out this concern, we offer two counteraguments. First, we plot pre-hearing trends in improper payments and use regressions to try to predict hearing incidence. Neither analysis suggests a relationship between pre-hearing changes in improper payments and hearings. Second, agencies count improper payments in the fiscal year after disbursement, at earliest, by sampling from the previous year's accounting database. This means that by the time Congress learns of rates and holds a hearing, the payment procedures of the agency might well have reformed or improved. Use of random samples also induces sampling variability into the process, further limiting tight endogeneity. To our read, these factors make a strong selection effect unlikely.

We offer two main findings. First, we find that congressional hearings do lead to subsequent declines in improper payments. When a congressional committee calls a witness from the bureaucracy to testify about improper payments, on average the rate of improper payments for that witness's agency declines in

subsequent fiscal years.

Second, while there is downward movement in rates following oversight, the magnitude of these declines are small relative to the scope of the problem. This is most clear when looking at the pattern of improper payments across the entire bureaucratic state, which despite decades of hearings and statutory reforms by Congress (along with attention from the executive), has shown limited improvement. Median estimated improper rates of programs directed to audit payments was 2.2% in 2005 and 2.7% in 2021 (75th percentiles 7.8% and 10.1%).

An additional result of our analysis is that program estimates of improper payments appear to increase in the first fiscal year following an oversight hearing. At first glance, this suggests that oversight has a counterproductive effect of increasing the problem that Congress sought to address through oversight. However, our reading of hearing transcripts suggests that this near-term increase is likely driven by agency officials increasing efforts to identify and measure improper payments following the hearing. This apparent increase in agency loss is more likely a more accurate measure of the problem than an actual increase. Declines in fiscal years beyond the first supports this interpretation.

To account for the possibility that oversight may happen outside of hearings (i.e., Selin and Moore, 2023), we also consider other techniques through which oversight of the executive branch could be conducted. We rule out a possible secondary channel of oversight, correspondence between individual legislators and agencies. We find that agency payments are rarely mentioned in these contacts, which Lowande (2018) finds to be largely driven by issues that have direct constituent or district interests. We also put the effectiveness of oversight hearings into the larger context of other actions taken by Congress and the president on improper payments. We evaluate the efficacy of appropriation committee reports, statutes, and executive actions on improper payments and find muted idiosyncratic effects that do not contain the systematic pattern seen with oversight hearings.

Our results suggest that Congress faces serious challenges in using oversight for political control. Discussion of oversight is frequently linked with other action that a legislature can take to influence the bureaucracy: statutory actions such as determining agency budgets, the amount of discretion granted to bureaucrats (e.g., Epstein and O'Halloran, 1994), and splitting authority across multiple agencies (Bils, 2020), or non-statutory actions such as specifying directives in committee reports (Bolton, 2022) or communicating the preferences of legislative actors (Lowande, 2018; Shipan, 2004). In the realm of improper payments, however, Congress has availed itself of many of these additional tools yet improper payments remain an as-yet

unsolved problem across the bureaucracy. Our unique measure of effectiveness, thus, allows us the conclusion that Congress's prospects for political control of the bureaucracy are in serious doubt.

1 Legislative efforts at control of the bureaucracy

Scholars disagree about how and whether Congress controls the bureaucracy. On the strong control side of the argument, the literature suggests that ex-post and ex-ante tools allow Congress to influence bureaucratic action. Scholars of the "congressional dominance" theories argued Congress controls the bureaucracy with ex post tools such as oversight and budgetary decision-making (Weingast and Moran, 1983; Weingast, 1984; McCubbins and Schwartz, 1984). Others suggest the power of ex ante controls: the legislature can design and use administrative procedures to control bureaucratic behavior (McCubbins, Noll, and Weingast, 1987; McCubbins, Noll, and Weingast, 1989) or delegate authority and discretion in a way that best ensures the implementation of the legislature's goals (Epstein and O'Halloran, 1994; Epstein and O'Halloran, 1999).

On the other side of the argument, critics such as Moe (1987) argue that theories of strong congressional control neglect the motivations, resources, and incentives of the bureaucracy, which might lead the bureaucracy to ignore or sidestep Congressional wishes. Even with control mechanisms such as oversight, appointments, budget, or new legislation, bureaucrats still possess their own autonomy such that Congress fails to achieve political control. Moe (1987) argues about congressional dominance theory, "While the theory presumes to explain congressional control of the bureaucracy, in fact it offers little more than an assertion supported by reference to legislative rewards and sanctions and their alleged efficacy in securing bureaucratic compliance (480)."

On both sides of the debate, oversight is commonly highlighted as a central mechanism available to Congress to influence the bureaucracy. McCubbins and Schwartz (1984) define oversight as the way in which Congress seeks to "detect and remedy executive-branch violations of legislative goals." They argue that a reelection-minded Congress prefers "fire alarm" oversight, with tools and procedures designed to alert Congress to violations that are most salient to interest groups and constituents, rather than "police patrol" oversight that calls for comprehensive, centralized monitoring. This view places Congress as engaging in oversight of the executive branch primarily to fulfill political goals, echoing other work where oversight is viewed as an activity used in legislators' pursuit of reelection (Ogul, 1976; Fenno, 1973).

Others step back from elections and suggest oversight helps stop "policy drift," where oversight occurs

when agencies take actions inconsistent with Congress's policy goals (Dodd and Schott, 1979; Kriner and Schwartz, 2008; McGrath, 2013). While policy goals are, of course, related to political goals, it remains unclear whether oversight is primarily driven by reelection motives or policy concerns.

Empirical work on oversight has largely focused on examining congressional committee hearings where the number (or presence of) oversight hearings conducted by congressional committees is used to measure oversight (e.g., Fowler, 2015). Scholars have found that congressional oversight activity has overall increased over time (MacDonald and McGrath, 2016; Aberbach, 1990), is driven by disagreement between congressional committees and the bureaucracy (Kriner and Schwartz, 2008; McGrath, 2013), and is more frequent during periods of divided government (Kriner and Schickler, 2016) or periods just after unified government is re-established as committees engage in "retrospective" oversight (MacDonald and McGrath, 2016).

Of course, congressional efforts at oversight are not limited to committee hearings. Congress conducts oversight through correspondence between legislators and bureaucrats (Ritchie, 2018; Lowande, 2018) and congressional participation in rulemaking (Lowande and Potter, 2021; Silfa, N.d.). Scholars have found that these interactions between legislators and bureaucrats are generally driven by constituency or district concerns rather than ideological patterns.

While these empirical studies have explained when and why Congress is likely to engage in oversight activities, they all stop at the point of oversight being conducted. While such analysis answers some important questions (e.g., if Congress engages in oversight when it is called for, as in McCubbins and Schwartz 1984 or MacDonald and McGrath 2016), it does not answer a question of central importance to our models of legislative control: *How much does congressional oversight influence bureaucratic behavior*? Much work assumes that oversight influences bureaucratic behavior and thus helps resolve its principal-agent problem. But, to date, we do not know how accurate this assumption and, therefore, how well oversight facilitates congressional control.

Research has circled the efficacy of oversight. Kriner and Schickler (2016) find that congressional oversight hearings can erode presidential approval ratings, giving presidents the incentive to change bureaucratic behavior. Oversight hearings also occur more often when they are most likely to have a policy effect (MacDonald and McGrath, 2016), implying that the purpose of hearings is to influence bureaucratic action. Studies of other nonstatutory means of control have found effects on bureaucratic behavior, such as agencies withdrawing policy proposals in response to the possibility of a legislative veto (Acs, 2019)

or agencies adjusting their own monitoring activities in response to shifting preferences of committees and Congress (Shipan, 2004). But no work, to our knowledge, directly estimates the empirical relationship between oversight hearings and magnitude of change in congressional agency loss to the bureaucracy.

One concern about oversight hearings as a method of congressional control is that they may be subject to forces unrelated to the putative goals of the oversight. Due to their public nature, the existence of a hearing or the content of the hearing itself may reflect ideological or partisan disagreement, reducing its potential effectiveness for true oversight. For instance, the House Select Committee on Benghazi oversight hearings, while ostensibly on the topic of conducting oversight and investigations into the attack on the U.S. consulate in Libya, were commonly interpreted as an effort to discredit Hillary Clinton as she pursued election to the presidency.

While oversight hearings in general can indeed be motivated by factors unrelated to the topic of the hearing—thus influencing the content of the hearing and potentially its effect on the bureaucracy—improper payments are, for better or worse, a topic that does not garner much publicity, political grandstanding, or partisan disagreement. As we show in the next section, reducing improper payments has been a goal of members and presidents of both parties, suggesting against oversight hearings on improper payments as political grandstanding.

2 Improper payments and federal bureaucracies

In order to measure the effectiveness of oversight, we require a measure of bureaucratic outcomes such that we can estimate the outcome with and without (or, before and after) oversight. This requirement is difficult to meet due to the nature of public policy. Government bureaucracies often produce public goods, which by their nature are difficult to value on some common scale and therefore difficult to compare across time and agencies.

We have identified one measure of bureaucratic outcomes that is, at least to first approximation, comparable across time and agencies. *Improper payments* are "any payment that should not have been made or that was made in an incorrect amount under statutory, contractual, administrative, or other legally applicable requirement" (Paymentaccuracy.gov, 2023). Payments made by the bureaucracy are denominated in dollars. Rate of improper payments made by federal agencies is a fraction that can be compared within agencies over time or between agencies. We argue that all members of the legislature (and likely the bureaucracy, as

well) agree that rate of improper payment should be minimized. Therefore, we have a common value target outcome on a common scale.

An additional benefit of this measure is that both Congress and the executive have worked to rein in improper payments made by federal bureaucracies for almost 25 years. Improper payments rose to the congressional agenda after the Government Management Reform Act of 1994 required major departments to prepare and have audited agencywide financial statements (Government Accountability Office, 1999, p. 5). The GAO identified $19.1 billion in improper payments when it reviewed nine audits from fiscal year 1998 and concluded, "Improper payments are much greater than have been disclosed (p. 6)". GAO made recommendations to the Office of Management and Budget (OMB) to improve internal controls and reduce improper payments.

The Senate held one hearing on improper payments in fiscal year 2000 with witnesses from the GAO and USAID. Congress's first statutory effort came with its August 2001 markup of H.R. 2586, the Defense Authorization Act for Fiscal Year 2002. The House Armed Services Committee inserted the "Erroneous Payments Recovery Act of 2001" as Section 811 to the thousand-plus page bill. Sections 812 and 813 directed executive agencies with contracts of total value greater than $500 million to conduct programs to identify and then recover erroneous payments made to federal contractors. Agencies would develop their own programs following guidelines to be created by OMB. The bill passed the House and went to conference with a Senate version without the provision. The conference report included the provision and became Public Law 107-107 with the signature of President George W. Bush on December 28, 2001.

The Improper Payments Information Act of 2002, signed into law November 26, 2002, provided further and more extensive guidance than the 2001 statute. The new law directed any executive program with estimated improper payments exceeding $10 million to audit and estimate improper payments, report on causes, state what infrastructure or information technology requirements would help reduce improper payments, and describe what steps had been taken to reduce such payments.

OMB issued guidance for agency implementation of the 2001 and 2002 laws in January and May of 2003 (OMB memoranda M-03-07 and M-03-13). OMB directed agencies to collect statistically-valid samples of payments to determine improper payment rates and report on the results to OMB and Congress effective fiscal year 2004. In a 2006 memorandum, OMB director Rob Portman updated and consolidated guidelines, allowing "low-risk" programs to audit payments once every three years rather than annually. High-risk programs would continue to implement annual audits and annual reports on efforts for remediation.

During the decade of the 2000s, Congress increased oversight on improper payments. Both the House and Senate held hearings with a total of 14 from fiscal 2003 through fiscal 2009.

On November 20, 2009, President Barack Obama issued Executive Order (EO) 13520 "Reducing Improper Payments." The EO directed implementation of new executive actions more punitive than previous OMB guidance. Agencies would have to publish annual information about improper payments, identify by name an agency official accountable for improper payments, and coordinate with other agencies to identify contractors and practices related to high improper payment rates.

OMB Memorandum M-10-13 implemented EO 13520 directives on March 22, 2010. Agencies would each have an "accountable official" to oversee and take responsibility for improper payments and who would issue an annual report to the agency's inspector general (IG). Programs would also have to publish target reductions in improper payment rates and publish the names of contractors who received improper payments but failed to return them in a timely manner. M-10-13 also set out the consequence, "If a high-priority program does not meet its supplemental targets for two consecutive years, then it is required to submit a report to the Director of OMB."

The Improper Payments Elimination and Recovery Act of 2010 (IPERA, July 22, 2010) codified many of the requirements from M-10-13 in statute. IPERA additionally required IGs to annually audit agencies to assure compliance with all requirements and reduced the threshold definition of "significant" overpayments. OMB guidance (M-11-16, April 14, 2011) reduced the threshold further from 2.5% to 1.5% improper payments as a percent of all payments, regardless of dollar amount, beginning in fiscal year 2014.

Congress held four, 15, and 11 hearings in fiscal years 2010, 2011, and 2012, with 80 witnesses from 13 different agencies called to testify. This oversight culminated in the Improper Payments Elimination and Recovery Improvement Act of 2012, signed into law January 10, 2013, which codified many of the requirements from M-11-16. It additionally required agencies to check contract awardees against databases of death records, excluded parties from the Government Services Agency, credit risks, and excluded individuals from Health and Human Services payments, along with any other databases identified by OMB. The "Do Not Pay Initiative" required all agencies to check all current and future awards against the databases by June 1, 2013.

From fiscal 2013 through 2021, Congress held 37 hearings with 93 witnesses from 12 different agencies. The Payment Integrity Information Act of 2019 (March 5, 2020) consolidated previous improper payment acts and allowed OMB to establish pilot programs to test potential accountability mechanisms for agencies.

The bill established an interagency working group on payment integrity. Issuing guidance M-21-19 (March 5, 2021) directed programs with annual outlays above $10 million to conduct risk assessment at least once every three years.

This history of congressional and executive efforts to reduce improper payments demonstrates the importance of the bureaucratic outcome. The setting provides a unique research opportunity because oversight hearings invited witnesses from different agencies in different fiscal years. This allows us to estimate the effect of oversight hearings on improper payment rates at the agency and program level by comparing rates of programs called to testify to rates of programs without hearings. In the next section we describe our data, research design, and strategy of identification.

3 Data, research design, and identification

To evaluate the effect of congressional oversight on improper payments in the bureaucracy, we compile and link data from federal fiscal years 2000-2021 on (1) improper payments by agency programs and (2) congressional committee hearings on improper payments.

3.1 Improper payments

We collect data on improper payments using a combination of government sources. First, the OMB-run website `PaymentAccuracy.gov` provides data on improper payments for federal fiscal years 2015 and forward. This data includes the improper payment amounts and rates at the program level for all programs subject to statutory reporting requirements.

To extend this improper payments data earlier than 2015, we collected annual financial reports for each agency. This required scouring current and archived versions of each agency's internet pages to locate a copy of the annual report. Each annual financial report includes improper payments estimates for at-risk programs in that agency. The starting year of improper payments data differs by agency based upon availability of archived financial reports. The earliest fiscal year for which coverage started was 2000 (Social Security Administration) and the latest fiscal year for which coverage started was 2012 (12 agencies).

Table A1 in the Appendix describes the year coverage by agency of our resulting improper payments dataset. Note that due to different years of reporting and at-risk determinations, different agencies and programs have different years of coverage.

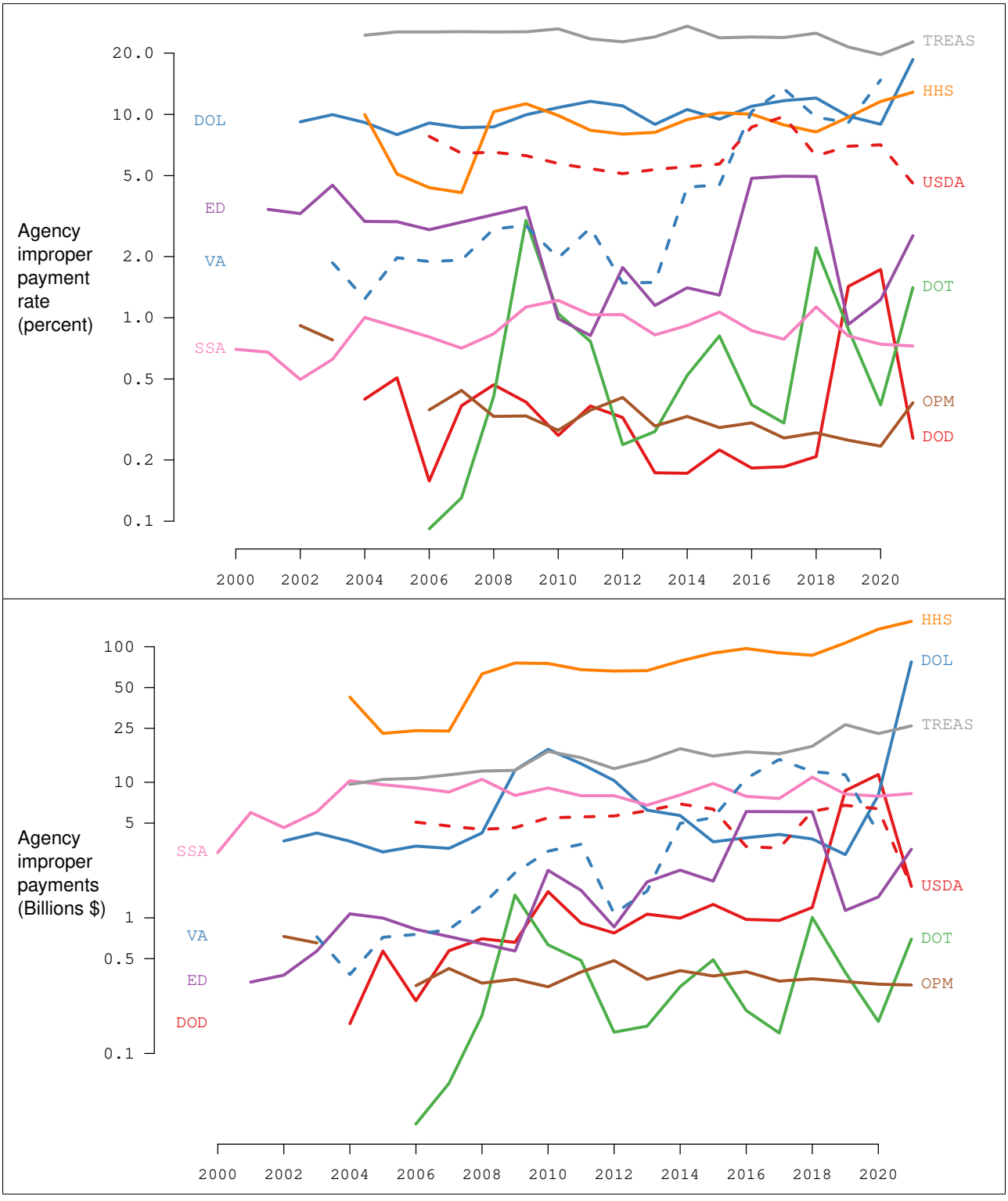
It is important to note that our improper payments data only covers programs subject to statutory reporting requirements from the legislation and OMB guidance discussed in the previous section. As a result, our analysis estimates the effect of oversight for programs deemed at-risk for substantial improper payments as defined by Congress and the executive branch. While this might not be the experiment one would design to estimate the most general effect of oversight, we believe the censored sample should be of limited concern. The potential effect of oversight on the set of programs deemed low-risk for improper payments seems likely to be lower than the potential effect on programs deemed at-risk. Low-risk programs have lower (or zero) improper payment rates, and so there is a ceiling on the efficacy of oversight on improving performance for the programs excluded from our sample. Averaging this smaller effect of low-risk programs excluded from our sample with a larger effect for at-risk included in our sample would yield an average effect of oversight on all programs smaller than what we estimate here. Because we conclude that the effectiveness of oversight is limited, the full-sample effect would likely demonstrate an even smaller effect of oversight, strengthening rather than weakening our conclusion.

In Figure 1 we present agency-level trends in improper payments by year for the ten largest-outlay agencies. The top frame presents improper payment rate as a percentage of total payments, the bottom the dollar amount of improper payments. The top frame shows that the Department of the Treasury had the highest rate of improper payments during this time period with more than one out of every five dollars in outlays deemed improper. This is because the IRS has a hard time confirming the accuracy of payments for the Earned Income Tax Credit program. Other agencies with large rates of improper payments are Labor, Agriculture, and Health and Human Services (HHS), whose improper payment rates are on the order of five to ten percent.

While some agencies have relatively constant rates of improper payments, others see year to year variability and secular trends. Our research design will ask whether congressional hearings predict these changes. As an overall picture, however, one would probably not conclude that Congress's two decades of effort to rein in improper payments have been successful. Typical improper payment rates in 2020 are no lower than typical rates in 2003. Our regression results confirm this observation, suggesting limits to Congress's ability to control the bureaucracy.

The bottom frame presents total improper payments by fiscal year and highlights the magnitude of the problem. In recent fiscal years, HHS has improper payments that exceed $100 billion, Labor above $50 billion, and Treasury above $10 billion. Somewhere around one half of improper payments are eventually

Figure 1: Improper payment rates and amounts by fiscal year and agency, ten largest agencies by total outlays. Y-axes on log scale.



recovered, so these numbers are larger than the total losses to the taxpayer. But even cutting these numbers in half indicates a large sum that one would imagine many members of Congress would like to use for other purposes.[2]

3.2 Oversight hearings

To identify congressional committee oversight hearings on improper payments, we use data from Ban, Park, and You (2023), which covers the names and affiliations of the witnesses testifying in those hearings, as well as the transcripts of these hearings. We identify hearings on improper payments as hearings which mention a set of phrases related to improper payments in at least one of four fields that represent the focus of a hearing: title, summary, subjects, or testimony subjects of the hearing, as specified by ProQuest Congressional.[3] By filtering for mentions of improper payments in these four descriptive fields of the hearings, we identify hearings that committees held with the intent of reviewing and discussing improper payments.[4]

Improper payment hearings invite testimony from employees of executive agencies and programs. We use the affiliations of the bureaucratic witnesses to identify the agencies overseen in the hearing by matching these witnesses to the 15 executive departments and 55 independent agencies in the federal government as defined by the Office of Personnel Management (we use the word "agency" to refer to the executive department or independent agency level).[5] Generally, these witnesses are Assistant Secretary or Director level within an agency or the Chief Financial Officer, Controller, Inspector General, or their deputies, of an agency or office within the agency. While hearings may mention certain programs or discuss program-level details, the witnesses invited have positions and job responsibilities at the *agency-level* and so we view improper payments hearings as agency-level stimulus. In Table 1 we summarize improper payment hearings held by Congress since 2000, tabulating the number of hearings, witnesses, and agencies called to oversight

[2]Recovery rate or amount recovered would be an outcome variable of interest. However, not until fiscal 2013 did statute require compilation of recovery numbers. More problematic is lack of direction as to when and how to pursue recovery or how to define the denominator for recovered funds. Agencies are directed to pursue recovery if they anticipate it to be "cost-effective," whose definition is left to the agency to determine. This means, in our view, the numbers are not comparable across agencies or even necessarily within-agency across time, unfortunately.

[3]The phrases we use are "improper payment," "payment integrity," "erroneous payment," "fraudulent payment," and "payment error."

[4]There are hearings in which a phrase about improper payments was mentioned in the full hearing text but not in the four fields representing the focus of the hearing. This occurred if a witness or legislator mentioned the issue off-hand. We exclude these hearings from our categorization of hearings on improper payments because they most likely do not reflect a committee conducting oversight on the issue.

[5]There are three agencies who received an improper payments hearing, but are not in the improper payments data from PaymentAccuracy.gov: Department of State, Nuclear Regulatory Commission, and the Securities and Exchange Commission. Additionally, employees from GAO and OMB appear in these hearings to testify on the guidance they have issued on reducing improper payments; since these two agencies are not the target of reducing such payments, we do not count them as treated agencies.

Table 1: Improper payments hearings by federal fiscal year. Witness counts exclude non-bureaucratic witnesses.

Fiscal year	Hearings	Agencies	Witnesses	Fiscal year	Hearings	Agencies	Witnesses
2000	1	2	3	2012	11	10	27
2003	3	5	8	2013	5	6	14
2004	1	2	2	2014	7	6	23
2005	2	4	6	2015	6	7	18
2006	3	8	15	2016	4	8	11
2007	2	6	8	2017	4	4	9
2008	1	1	1	2018	4	4	8
2009	1	3	4	2019	2	3	4
2010	4	4	11	2020	2	3	3
2011	15	11	42	2021	3	2	3

by fiscal year (summing across House and Senate).

3.3 Research design and Identification

To estimate the effect of improper payment hearings on improper payment rates, we use a two-way fixed effects identification strategy. Because we are uncertain about how long it might take for agencies to enact reforms to improve their improper payment problems, we estimate a linear trend in improper payments in the years following a hearing. Our specification is:

$$\{\text{Improper payment rate}\}_{it} = \alpha_i + \delta_t + \beta \text{Hearing}_{it} + \gamma\{\text{Years since last hearing}\}_{it} + \varepsilon_{it} \quad (1)$$

where $\{\text{Improper payment rate}\}_{it}$ measures the improper payment rate of agency i in fiscal year t, Hearing_{it} takes the value one if a witness from agency i testified in at least one oversight hearing on improper payments in year t, $\{\text{Years since last hearing}\}_{it}$ counts the number of fiscal years since a witness from that agency was most recently called to testify (year of hearing equals zero), and α and δ are agency and fiscal year fixed effects. These fixed effects capture time-invariant features of agencies (policy domain, relative budget size, etc.) and factors that affect all agencies in each fiscal year (new statutes or executive orders, budget environment, economic conditions, focus of presidential administration, GAO involvement in hearings, etc.).

The coefficient β measures the effect of oversight hearing on improper payments in the year of the

hearing.[6] Because agencies estimate improper payments by sampling from payment databases in fiscal years subsequent to that of the payments, the hearing might influence the measurement of IP rate in the fiscal year of the hearing. To estimate improper payments in fiscal year 2010, an agency would take a sample from their fiscal 2010 database of payments in fiscal 2011 and then evaluate the validity of each sampled 2010 payment. Thus a hearing in fiscal 2010 could influence either the payment choices in the remainder of fiscal 2010 or the evaluation efforts in fiscal 2011. Because of the multiple causal pathways influencing β, this coefficient might be positive if the agency exerts additional efforts to identify improper payments following the hearing, negative if the agency exerts additional efforts to prevent improper payments following the hearing, or some mixture of the two.

Due to the ambiguous predictions about β, we also estimate the time-forward linear effect of the hearing with the coefficient γ. If oversight leads to improvement in the agency's payment procedures, improper payment rate should improve (negative coefficient) in the years following the hearing. Thus γ serves as our key estimate of the efficacy of congressional oversight. If oversight is effective, γ should be substantively important and negative.

We complement an agency-level specification above with a program-level regression with the same specification but for i indexing programs instead of agencies. Hearings are measured at the agency level because, in our review, nearly all witnesses represent agencies (e.g., inspectors general, payment administrators for the full agency) rather than individual programs. We therefore cluster standard errors on the agency-year for program-level analysis.

Unbiased estimates of β and γ require the usual two-way fixed effects assumption of parallel trends, that the path of improper payment rate for agencies called to testify would have followed the path of agencies not called to testify were the witness not invited to the hearing. In the Appendix, we offer two evaluations of the parallel trends assumption. In Figure A1, we plot time-series of improper payment rates by program for programs with witnesses called to testify. The figure demonstrates no pattern of improper payment rates in the years prior to a hearing. In Table A2, we regress an indicator for oversight hearing in that fiscal year on lags of proportional change in IP rate and on total dollar amount of improper payments. We find that changes in improper payment rates do not predict incidence of oversight hearings.

[6] An additional specification of interest would be to test whether there is an effect of *any* kind of hearing on an agency (not just hearings on improper payments), to see if there are any spillover effects from any hearing in general on an agency's improper payments. However, at least one witness testified before Congress in every fiscal year from all but one of the agencies in our data set (the Corporation for National and Community Service [AmeriCorps]).

4 Legislative control and oversight results

In Table 2, we present estimates from specification (1) at the agency (columns one through three) and program (four through six) level. Column one presents results combining House and Senate hearings into single variables. The coefficients indicate that in each year following a hearing with a witness from that agency, improper payment rate falls by about 0.1 percentage points.[7]

Table 2: Dynamic Effect of Hearings on Improper Payment Rate. Dependent variable is improper payment rate as a percent of all payments.

	Agency-level			Program-level		
	(1)	(2)	(3)	(4)	(5)	(6)
Years Since Last Agency Hearing	-0.10***	-0.082**		-0.17***	-0.17***	
	(0.035)	(0.040)		(0.053)	(0.063)	
Years Since Last Unified Govt Hearing		-0.082			0.12	
		(0.074)			(0.14)	
Years Since Last House Agency Hearing			-0.079**			-0.059
			(0.032)			(0.041)
Years Since Last Senate Agency Hearing			-0.10			-0.27**
			(0.062)			(0.11)
Agency Hearing	0.58	0.62		1.34	0.19	
	(0.49)	(0.57)		(1.03)	(1.10)	
Agency Hearing * Unified Government		-0.32			3.79	
		(0.79)			(2.27)	
House Agency Hearing			0.87			1.32
			(0.54)			(1.12)
Senate Agency Hearing			-0.41			1.12
			(0.62)			(0.74)
Fiscal Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Agency/Program FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	302	302	302	1284	1150	1284

Standard errors in parentheses; clustered on agency-year for program-level analysis.

** $p < 0.05$, *** $p < 0.01$

In column two we estimate variability in effects by whether or not the hearing was held under unified government. While the standard error is large, the point estimate suggests that hearings held under unified government lead to twice the decline in IP rate than hearings held under divided government. Column three

[7] In the year of the hearing, the agency's IP rate is about 0.6 percentage points higher, all else equal, but the standard error makes this estimate highly uncertain. A few agencies had hearings on erroneous payments decades ago. We top-code the years since hearing variable at 20 years on the assumption that the effect of hearings held more than 20 years ago is unlikely to be much greater than zero.

suggests that both House and Senate hearings lead to subsequent declines in improper payments. The two coefficients are -0.08 (House) and -0.1 (Senate) with standard errors that do not allow us to distinguish between them.

Columns four through six present the same specifications for a program-level analysis. We find that improper payment rates decline, on average, by about 0.17 percentage points each year following a hearing (column four). We do not find evidence that this decline varies by whether or not the hearing happened under unified government (column five). Point estimates suggest that Senate hearings are more consequential than House hearings at the program level (column six).

The results of Table 2 suggest that oversight hearings lead to a subsequent decline in improper payment rates for agencies with witnesses called to testify. Oversight is effective. However, the magnitude of the decline is somewhat limited. For each year beyond the hearing, improper payment rate improves by 0.1 (agency) to 0.2 (program) percent, on average.

4.1 Moderators of legislative control

We next consider whether characteristics of the hearing moderate the effect on improper payments. We collected four hearing characteristics that each try to approximate higher quality of hearing or higher stimulus to the agency: (1) the number of agency witnesses called to testify across all hearings on improper payments in that fiscal year, (2) the length of hearings with witnesses from that agency measured by log word count, (3) the number of total witnesses (i.e., including non-agency witnesses such as those from GAO or OMB in the count) called to testify in improper payment hearings referencing that agency in that fiscal year, and (4) the number of agency witnesses who are political appointees called to testify in that fiscal year. We also collect two additional hearing characteristics that capture higher exposure of the hearing for the agency: (5) whether the hearing was at the full committee or subcommittee level and (6) whether the hearing was held during a presidential election year. While none of these variables perfectly measure the "strength of oversight" we would like to capture, they each circle that target. The first two moderators are perhaps the closest proxies, as the number of agency witnesses called to testify and the length of the hearing are directly increasing with amount of attention and detail the committee gives to the issue. Table 3 presents the results on the dynamic effect of the first two moderators on improper payment rates; Appendix Table A3 presents the results for the last four moderators.

None of the moderators in Table 3 offer clear improvement in fit. Standard errors are generally large.

Table 3: Dynamic Effect of Moderators on Improper Payment Rate. Dependent variable is improper payment rate as a percent of all payments.

	Agency-level		Program-level	
	(1)	(2)	(3)	(4)
Years Since Last Agency Hearing	-0.087**	-0.11	-0.18***	0.090
	(0.038)	(0.13)	(0.055)	(0.18)
Years Since Last * Length of Hearing Transcripts, Fiscal Year of Last	-0.0089		0.0068	
	(0.0078)		(0.013)	
Years Since Last * Number Agency Witnesses, Fiscal Year of Last		0.0020		-0.26
		(0.13)		(0.16)
Agency Hearing	-2.17	0.48	1.63	0.93
	(1.79)	(0.65)	(2.10)	(1.03)
Length of Hearing Transcripts (log words), This Fiscal Year	0.28		-0.020	
	(0.18)		(0.25)	
Number of Agency Witnesses, This Fiscal Year		0.055		0.19
		(0.28)		(0.26)
Constant	5.70***	5.55***	5.93***	6.25***
	(0.32)	(0.30)	(0.53)	(0.46)
Fiscal Year FE	Yes	Yes	Yes	Yes
Agency/Program FE	Yes	Yes	Yes	Yes
Observations	249	249	1150	1150

Standard errors in parentheses; clustered on agency-year for program-level analysis.

** $p < 0.05$, *** $p < 0.01$

Point estimates offer some suggestive evidence at the agency level that the length of the hearing leads to an initial increase in estimated improper payment rate in the first year following the hearing but then leads to a decline in years after that. This is similar to our main finding above but with larger sampling uncertainty. Corresponding point estimates at the program level, however, run counter to this pattern. Point estimates for the number of agency witnesses follows the pattern of the main finding at the program level, but is negligible with large standard errors at the agency level.[8]

Large standard errors are also found for the number of total witnesses and whether the hearing was held at the subcommittee or full committee level in Table A3. We find some evidence that a hearing called with more agency witnesses who are political appointees leads to increases in improper payment rates at the agency level. Table A3 also suggests at the agency level that a hearing held during a presidential general election year leads to declines in the improper payment rate in succeeding years.[9] This could be explained by committees changing the quality or selection of their oversight, agencies changing their responsiveness, or both. Our model does not allow us to distinguish between those possibilities. However, other work on investigative hearings finds that the electoral calendar does not lead to higher oversight activity by committees (Kriner and Schickler, 2016). If committees are similarly not changing their behavior with improper payment hearings in the lead up to a presidential election, then agencies may be increasing responsiveness and implementing changes that have long-lasting effects during these periods.

5 Other methods of legislative and executive control

Our main analysis centers on congressional committee hearings as the tool of oversight, following existing literature that focuses on hearings (e.g., Aberbach, 1990; Fowler, 2015; MacDonald and McGrath, 2016; Kriner and Schickler, 2016; McGrath, 2013). Might congressional oversight of improper payments, however, operate through other channels?

Congress, its committees, or individual members may use non-hearing techniques to conduct oversight depending on the political context or purpose (Selin and Moore, 2023). Congress can use non-oversight methods to influence bureaucratic action. The President also uses various actions to mitigate the principal-

[8]Large standard errors are also found for the number of total witnesses and number of agency witnesses who are political appointees in Table A3.

[9]This effect does not seem to be driven by 2012, a year in which Congress held 11 hearings on improper payments and formed the Improper Payments Elimination and Recovery Improvement Act of 2012. Column (5) in Table A3 drops hearings from fiscal year 2012.

agent problem within the executive branch. In this section, we consider the use of these other methods of legislative and executive control to reduce improper payments.

We begin with the congressional practice of correspondence as a method of control. Studies such as Lowande (2018) and Ritchie (2018) focus on inquiries (contacts) made by individual legislators to agencies. This form of communication is informal and nonstatutory but, as with hearings, can take the form of oversight as communication between legislature and executive about legislative priorities.

Using data from Lowande and Ritchie (2020), we examine all contacts between legislators and agencies from the time period available that overlaps with our main analysis, fiscal years 2005-2014. We find that out of the 11 agencies for which contacts data are available for this time period, the description for 80 out of 34,419 total contacts refers to a topic related to agency payment issues.[10] This small number of contacts, less than our total count of hearings, implies that while contacts between individual legislators and agencies could be a channel of oversight for some matters, it is not the standard method for improper payments.[11]

5.1 Comparison with appropriations committee reports, statutes, and executive actions

The principals of the bureaucracy—Congress and the president—have methods outside of what is typically considered oversight to try and control agency behavior. Congress can legislate and issue appropriation committee reports. The president can issue executive orders. OMB can issue memoranda and guidance. While these sets of actions are not typically considered oversight, they are tools to motivate action within the bureaucracy. We next compare the effectiveness of oversight to the effectiveness of congressional and executive tools that could also influence improper payments.[12]

Scholars have characterized committee reports issued during the appropriations process as a channel through which the committees can instruct agencies on bureaucratic priorities, reporting requirements, or requirements to conduct certain activities or investigations (Bolton and Thrower, 2019; Kiewiet and McCubbins, 1991). Bolton (2022) and Schick (2008) argue that the appropriations committees can indirectly sustain agency cooperation through language in their annual reports.

We examine whether committee reports from the House Appropriations Committee addressing improper

[10] We use the phrases from our dictionary used to identify hearings on improper payments plus phrases "payment issue," "underpayment," and "overpayment."

[11] The low frequency of contacts about improper payments is in line with findings that the correspondence between individual legislators and agencies are largely driven by oversight issues that have constituent or district interests (Lowande, 2018); improper payments are a bureaucratic issue with little-to-no constituent or district concern.

[12] We note that because statutes, executive orders, and OMB memoranda apply to all agencies, rather than the agency-year level as with hearings and appropriation committee reports, there is less variation and so a larger inferential challenge for these actions.

payments reduce subsequent improper payment rates. We collect the text of all committee reports issued by the House Appropriations Committee in fiscal years 2000 to 2021 from Congress.gov. We classify committee reports as addressing improper payments if the report text mention one or more of the phrases relating to improper payments.[13] This process yields 75 appropriations committee reports addressing improper payments.[14]

Appropriation committee reports addressing improper payments can cover agencies in two ways. The report can cover an agency and directly refer to the agency's own improper payments as needing attention. Or, the report can cover an agency but not directly refer to an agency's improper payments, instead mentioning improper payments in general or mentioning a different agency's improper payments. (See Appendix Section C for an example of each case.) We manually code both of these instances for each report.[15]

Appendix Table A4 presents the results on the effect of appropriations committee reports addressing improper payments on improper payment rates. Columns one and two, (agency-level) three and four (program-level), show the estimated effect of an agency or program's agency having its improper payments directly mentioned in a report. Point estimates are negative for both agency-level and program-level improper payment rates, though with wide sampling variability in both cases. The point estimates do suggest some influence of appropriation reports, with agency-level declines of about 0.2 points and program-level declines of more than one percentage point. The program-level decline is statistically significant in column five but not six. We evaluate this evidence as suggestive if not definitive that appropriation reports can influence bureaucratic performance.

In Figure 2 we evaluate evidence on the efficacy of executive actions (the Obama executive order and significant OMB memoranda) and enacted statutes. The top frame plots the government-wide improper payment rate for fiscal years 2005 to 2021. The solid line is the mean rate across agencies, the dashed line a mean weighted by outlay amounts. We plot vertical lines at four fiscal years with salient executive or statutory actions, listed in the figure note.

The fiscal year trends suggest some evidence of influence following the fiscal 2010 actions. Notably, the executive order and OMB guidance occur early enough in the fiscal year (November 2009 and March 2010)

[13] We use the same list of phrases used to identify hearings on improper payments: "improper payment," "payment integrity," "erroneous payment," "fraudulent payment," and "payment error."

[14] These reports were not exclusively on the topic of improper payments, largely detailing the annual appropriations for agencies, but included at least one section or subsection specifically addressing improper payments.

[15] These independent variables are coded at the agency level. While there are times in which the report will mention the improper payments of a particular program of an agency, this happens rarely; when this does happen, the name of the agency is also mentioned and we code this as an agency-level mention for consistency.

to influence payment practices for much of fiscal 2010. Both dollar-weighted and raw means decline from fiscal 2009 to fiscal 2010 and on to fiscal 2013 where they bottom out before beginning to rise after fiscal 2016. That said, we see no evidence of declines following the 2013, 2014, or 2020 executive actions and statutes, and the decline following 2009 is from an improper payment rate of around 5 percent to a rate of around 3.75 percent in 2013.

The aggregate means do not account for changes in composition of the programs reporting improper payments, which is a function of each program's risk profile and mandatory evaluation schedule. The bottom frame attempts to fix the estimates by plotting the fiscal year fixed effects (δ_t in Eq. 1) from the models in Table 1, which include agency or program fixed effects (α_i in Eq. 1) to account for compositional changes. Above each fiscal year on the x-axis we plot the fixed effect estimate from the agency-level model (solid circle) and from the program-level model (solid square) along with 95% confidence intervals. The point estimates from both models replicate the pattern from the top frame, with the pre-2010 fixed effects greater than the 2010 through 2014 fixed effects followed by an increase towards 2020.

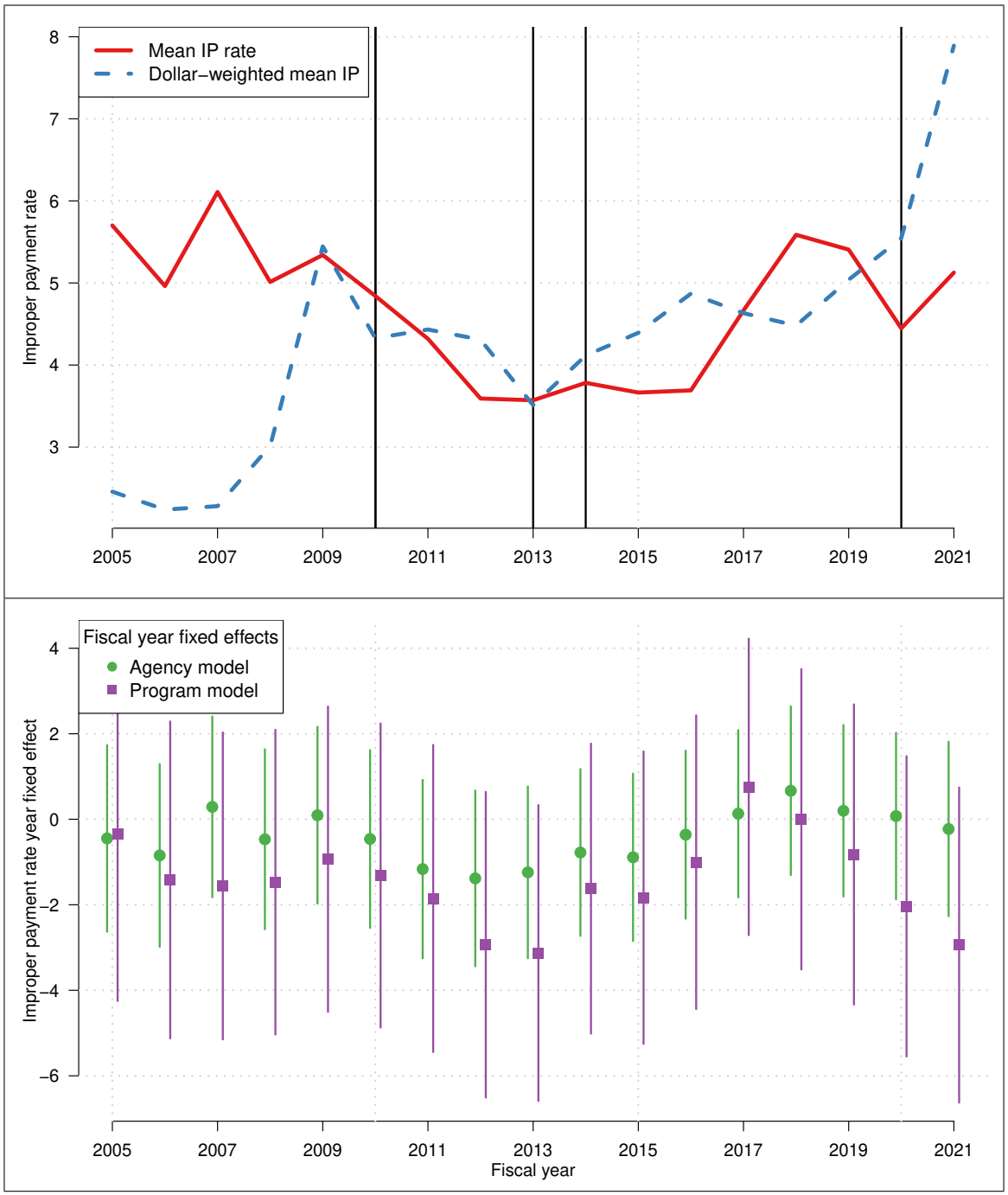
The point estimates suggest a decline in improper payment rates following fiscal 2018, particularly when accounting for program-specific factors in the program model. The size of the confidence intervals, however, does not allow for strong inferences about this pattern. Data from future years will allow evaluation of the consistency of this decline.

Overall, while there are some years in which non-oversight congressional and executive actions may have reduced improper payments, the effectiveness of these actions have been muted and inconsistent. The magnitude of the change in improper payments is, at most, 1.25 percentage points over a specific four year time period (fiscal years 2009 to 2013), which is not reproduced in years following. Overall, statutes, executive orders, and OMB guidance appear to offer, at best, the same limited efficacy of oversight hearings.

6 What works?

How exactly do agencies improve payment integrity? We consider two large programs who have improved their improper payment rates during the period of our study, Medicare Part C (Medicare Advantage) and Medicare Part D (Drug Benefit). Medicare Part C's average improper payment rate declined from around 13 percent in 2008 to around 8 percent in 2021 and Part D's from 3.25 percent in 2011 to 1.25 percent in 2021 (the years each program reports improper payments, see Figure A2).

Figure 2: Improper payment rate trends and model fixed effect estimates by fiscal year.



Note: Lines in lower frame extend to 95% confidence intervals. Vertical lines match to the following events: Executive Order 13520, OMB memorandum M-10-13, and IPERA 2010 statute (fiscal 2010); IPERIA 2012 statute (fiscal 2013); threshold defining significant overpayments reduced to 1.5% (fiscal 2014); PIIA 2019 statute (fiscal 2020).

Congress requires that each agency's annual financial report (AFR) document a Corrective Action Plan (CAP) stating the actions to be taken by at-risk programs to improve payment integrity. We reviewed the Health and Human Services (HHS) AFRs (Department of Health and Human Services, 2023) to identify the actions documented by each program. The 2008 Medicare Part C CAP described two actions. First, monthly payment validation by the agency. Second, national training sessions for Medicare Advantage Plan participants on effective documentation of risk adjustment factors and medical board reviews. The 2010 CAP repeated the 2008 actions and added monthly user group calls along with development of a new study to identify the factors associated with errors that would then be explained in outreach to providers. The 2012 CAP noted that the majority of payment errors were caused by insufficient documentation by plan providers. Corrective actions included those presented in 2008 and 2010 plus contract-level audits and physician outreach. The 2014 CAP reported new regulations requiring providers to report and return overpayments identified by the provider along with a payment recovery and appeal process. The 2016 CAP noted that organizations returned $317 million in self-reported overpayments and also reported having posted a request for proposal to hire a "Recovery Audit Contractor" (unsuccessfully filled). The 2018 CAP noted that organizations returned $65 million in self-reported overpayments and reported moving away from large webinar to smaller, in-person training for participating plans. The 2020 CAP reported continuation of previous efforts.

Medicare Part D's 2012 CAP included monthly payment validation in the payment system, national training programs, and plans for "in-depth analyses" to understand improper payments. The 2014 CAP noted continued national training programs, outreach to plan sponsors with invalid documentation, and the regulatory provisions also noted in the Part C plan. The 2016 CAP noted organizations returned $9.5 million in self-reported overpayments and stated continuation of current corrective action. The 2018 CAP noted return of $2.1 million in overpayments, a program to distribute Plan Sponsor Summary Reports to plans with history of invalid documentation, and the same change in training as the Part C report. The 2020 CAP reported action on outreach and training only.

The CAPs do not make clear what exactly allowed these two programs to improve their rates of improper payments. But it does appear that vendor education and outreach, to encourage appropriate documentation and to clarify processes and procedures, were central to the efforts of both.

7 Conclusion

Congress holds the power and responsibility to conduct oversight of the executive branch to produce effective democratic government. This oversight role has spawned many theories of congressional influence on the bureaucracy. On one end, theories suggest Congress controls the bureaucracy and can successfully address problems with "fire alarm" oversight (Weingast and Moran, 1983; Weingast, 1984; Calvert and Weingast, 1987; Moran and Weingast, 1982; Fiorina, 1981; McCubbins and Schwartz, 1984). On the other end, critics of dominance theories point out that congressional control of the bureaucracy is suspect for many reasons. Moe (1987) reminds the literature that bureaucrats have their own motivations and incentives that do not make them always subservient to Congress, and questions the efficacy of the tools practically available to Congress to influence the bureaucracy.

Empirical scholars have worked to assess congressional oversight. The majority of this work, however, has focused on when and with how much effort Congress conducts oversight. Oversight is primarily measured by the number of congressional hearings (Aberbach, 1990; Fowler, 2015; MacDonald and McGrath, 2016; Kriner and Schickler, 2016; McGrath, 2013). This literature, though, has remained squarely on the occurrence of oversight, with only some case studies on the consequences of oversight. Understanding the effectiveness of oversight activities is critical to evaluating the congressional check on executive action.

We provide a novel empirical study of the efficacy of congressional oversight on an outcome that can be measured across agencies and over time. By examining improper payments, an outcome for which there is near-universal support within Congress to reduce, we can evaluate how much congressional oversight influences bureaucratic behavior.

The answer appears to be that oversight is quite circumscribed. We find that oversight hearings on improper payments first increase attention to the problem and then lead to subsequent declines in improper payments. However, these declines are quite modest relative to the scope of the problem, around 0.1 percentage points per year for an agency with witnesses called to testify. We do not find that a second tool of oversight, correspondence between legislators and agencies, is used much at all. Statutes codifying reporting and recovery requirements and executive actions including OMB guidance to agencies or executive orders produce similar limited muted effects.

The improper payment setting seems like an opportunistic scenario for effective oversight. The goal of reducing improper payments is bipartisan and shared by individuals in both branches of government.

That we find such limited effect of oversight here begs the question of how oversight proceeds when the executive's goals diverge from the legislature or when partisan disagreement in Congress yields mixed messaging to the bureaucracy. On the other hand, perhaps oversight is most effective when out-party members of Congress use their oversight platform to embarrass the bureaucratic efforts of the other party's president. We leave this open question to future research.

In finding that oversight at first generates a so-called Hawthorne effect – inducing additional attention to improper payments by agencies subjected to oversight – our contribution provides new insights on both the limitations and prospects for oversight. Oversight hearings do not seem to trigger substantial near-term mitigation or correction of bureaucratic action from the perspective of Congress. Instead, oversight leads to increased attention to the bureaucratic deficiency and, ultimately, modest remediation of the problem.

Our findings have limitations and implications that provide a path toward future work. First, our study focuses on oversight hearings as Congress's main tool for oversight, following previous work. Congressional scholars have recently worked to consider non-hearing techniques used for oversight (e.g., Selin and Moore, 2023); our findings on the ineffectiveness of oversight hearings suggests the importance of gaining a full accounting of congressional oversight activities. While we also examine the non-hearing techniques of correspondence between legislators and agencies, appropriation committee reports, and executive actions, we have still been unable to uncover strong evidence of either legislative or executive control over bureaucratic deficiency. Additional work on other oversight activities can help us understand if other legislative or executive actions better motivate bureaucratic action.

Second, a critical interpretation of our result that improper payment rates decline in the few years after a hearing but do not globally improve across the bureaucracy is that oversight generates short-term attention by the agency. As time passes, however, the agency subject to oversight stops paying attention to the issue. The timeframe and sampling variability of our study make this interpretation difficult to evaluate. Were it the case, it would imply oversight of even smaller effect that what we conclude.

Finally, our study takes advantage of improper payments as a quantitative and comparable measure across agencies and time, as well as an issue that members of Congress from both parties have tried to resolve for decades. One possible drawback of improper payments, however, is that they are not as electorally salient for members of Congress as other issues. Members may choose to exert effective oversight on issues for which they can credit claim in a particularistic or partisan way. On the other hand, the fact that improper payments are a purely bureaucratic issue that is relatively insulated from electoral politics makes it an ideal

outcome to test the effects of *genuine* oversight. Oversight efforts on technical bureaucratic issues are much more likely to reflect true oversight efforts, rather than be a guise for grandstanding, position-taking, or other partisan rhetoric that do not represent actual efforts to change bureaucratic behavior. Future work should examine other issues and incorporate member incentives to probe whether effective oversight is driven by electoral concerns, while continuing to carefully disentangle true oversight efforts from when it is used as mere pretense.

Nevertheless, our findings suggest stark limits to Congress's best known tool for oversight on a standard bureaucratic issue. This new understanding of oversight is especially crucial when viewed in light of the enormous size of agency programs that touch many facets of Americans' lives. Medicare Part C by itself had improper payments of $23 billion in fiscal year 2021. The National School Lunch Program had $723 million and the Federal Pell Grant Program $521 million. In total, the GAO estimates that federal improper payments in fiscal year 2022 of $247 billion with a cumulative total of $2.4 trillion since fiscal 2003 (Government Accountability Office, 2023). These numbers could understate the total problem as the GAO found that only half of agencies were fully compliant with improper payment reporting requirements in fiscal 2021 (Government Accountability Office, 2023).

While investing in congressional capacity (see the programs studied in Fong, Lowande, and Rauh, 2023) might foster more and more effective oversight, oversight on improper payments seems mostly to be an information-gathering and information-production exercise. Congress and OMB seem to have plenty of information about the size and scope of the problem. It is not clear that more information generated by more effective oversight would be the solution to this challenge.

Perhaps substantial sums of payments improperly made are just the "cost of doing business" for a large government pursuing complex societal-level goals. It is certainly a large cost and, unfortunately, congressional oversight does not appear to provide much relief, offering only limited political control over bureaucratic deficiencies.

References

Aberbach, Joel D. (1990). *Keeping a Watchful Eye: The Politics of Congressional Oversight*. Washington, DC: Brookings Institution.

Acs, Alex (2019). "Congress and Administrative Policymaking: Identifying Congressional Veto Power". *American Journal of Political Science* 63.3, pp. 513–529.

Ban, Pamela, Ju Yeon Park, and Hye Young You (2023). "How Are Politicians Informed? Witnesses and Information Provision in Congress". *American Political Science Review* 117.1, pp. 122–139.

Bils, Peter (2020). "Policymaking with Multiple Agencies". *American Journal of Political Science* 64.3, pp. 634–648.

Bolton, Alexander (2022). "Gridlock, Bureaucratic Control, and Non-Statutory Policymaking in Congress". *American Journal of Political Science* 66.1, pp. 238–254.

Bolton, Alexander and Sharece Thrower (2019). "The Constraining Power of the Purse: Executive Discretion and Legislative Appropriations". *Journal of Politics* 81.4, 1266–1281.

Calvert Randall L., Mark Moran and Barry R. Weingast (1987). "Congressional Influence Over Policymaking: The Case of the FTC". In: *Theories of Congress: The New Institutionalism*. Cambridge University Press.

Clinton, Joshua D., David E. Lewis, and Jennifer L. Selin (2014). "Influencing the Bureaucracy: The Irony of Congressional Oversight". *American Journal of Political Science* 58.2, pp. 387–401.

Department of Health and Human Services (2023). *HHS Agency Financial Reports (AFR)*. URL: www.hhs.gov/about/agencies/asfr/finance/financial-policy-library/agency-financial-reports/index.html. Retrieved 8 May 2023.

Dodd, LC and RL Schott (1979). *Congress and the administrative state*. New York: Wiley.

Epstein, David and Sharyn O'Halloran (1994). "Administrative Procedures, Information, and Agency Discretion". *American Journal of Political Science* 38.3, pp. 697–722.

— (1999). *Delegating Powers: A Transaction Cost Politics Approach to Policy Making under Separate Powers*. Cambridge University Press.

Fenno, Richard F. (1973). *Congressmen in Committees*. Little, Brown & Co.

Fiorina, Morris P. (1981). "Congressional Control of the Bureaucracy: A Mismatch of Incentives and Capabilities". In: *Congress Reconsidered 2nd ed.* Congressional Quarterly Press.

Fong, Christian, Kenneth Lowande, and Adam Rauh (2023). *Expertise Acquisition in Congress*. Working paper, University of Michigan; Retrieved 12 July 2023. URL: https://lowande.polisci.lsa.umich.edu.

Fowler, Linda L. (2015). *Watchdogs on the Hill: The Decline of Congressional Oversight of U.S. Foreign Relations*. Princeton University Press.

Gilmour, John B. and David E. Lewis (2006). "Does Performance Budgeting Work? An Examination of the Office of Management and Budget's PART Scores". *Public Administrative Review* 66.5, pp. 742–752.

Government Accountability Office (1999). *Managing for Results: Strengthening Regulatory Agencies' Performance Management Practices*. URL: www.gao.gov/products/ggd-00-10. Retrieved 3 March 2023.

— (2022). *F-35 Joint Strike Fighter: Cost Growth and Schedule Delays Continue*. URL: www.gao.gov/products/gao-22-105128. Retrieved 3 March 2023.

— (2023). *Fiscal Year 2022 Estimates and Opportunities for Improvement*. URL: www.gao.gov/products/gao-23-106285. Retrieved 19 July 2023.

Kiewiet, Rodreick D. and Mathew D. McCubbins (1991). *The Logic of Delegation: Congressional Parties and the Appropriations Process*. University of Chicago Press.

Kriner, Douglas and Eric Schickler (2016). *Investigating the President*. Princeton University Press.

Kriner, Douglas and Liam Schwartz (2008). "Divided Government and Congressional Investigations". *Legislative Studies Quarterly* 33.2, pp. 295–321.

Lewis, David E. (2007). "Testing Pendleton's Premise: Do Political Appointees Make Worse Bureaucrats?" *Journal of Politics* 69.4, pp. 1073–1088.

Lowande, Kenneth (2018). "Who Polices the Administrative State?" *American Political Science Review* 112.4, pp. 874–890.

Lowande, Kenneth and Justin Peck (2017). "Congressional Investigations and the Electoral Connection". *Journal of Law, Economics, & Organization* 33.1, pp. 1–27.

Lowande, Kenneth and Rachel Augustine Potter (2021). "Congressional Oversight Revisited: Politics and Procedure in Agency Rulemaking". *Journal of Politics* 83.1, pp. 401–408.

Lowande, Kenneth and Melinda N. Ritchie (2020). *Congress Called*. Dataset retrieved 22 May 2023. URL: https://congresscalled.org.

MacDonald, Jason and Robert McGrath (2016). "Retrospective Congressional Oversight and the Dynamics of Legislative Influence over the Bureaucracy". *Legislative Studies Quarterly* 41.4, pp. 899–934.

McCubbins, Mathew D., Roger G. Noll, and Barry R. Weingast (1987). "Administrative Procedures as Instruments of Political Control". *Journal of Law, Economics, and Organization* 3.2, pp. 243–277.

— (1989). "Structure and Process, Politics and Policy: Administrative Arrangements and the Political Control of Agencies". *Virginia Law Review* 75.2, pp. 431–482.

McCubbins, Mathew D. and Thomas Schwartz (1984). "Congressional Oversight Overlooked: Police Patrols versus Fire Alarms". *American Journal of Political Science* 28.1, pp. 165–179.

McGrath, Robert J. (2013). "Congressional Oversight Hearings and Policy Control". *Legislative Studies Quarterly* 38.3, pp. 349–376.

Moe, Terry M. (1987). "An Assessment of the Positive Theory of 'Congressional Dominance'". *Legislative Studies Quarterly* 12.4, pp. 475–520.

Moran, Mark J. and Barry R. Weingast (1982). "Congress as the Source of Regulatory Decisions: The Case of the Federal Trade Commission". *American Economic Review* 72, pp. 109–113.

Ogul, Morris S. (1976). *Congress Oversees the Bureaucracy: Studies in Legislative Supervision*. University of Pittsburgh Press.

Paymentaccuracy.gov (2023). *About Paymentaccuracy.gov*. Office of Management and Budget. URL: www.paymentaccuracy.gov/about-payment-accuracy/. Retrieved 28 February 2023.

Ritchie, Melinda N. (2018). "Back-Channel Representation: A Study of the Strategic Communication of Senators with the US Department of Labor". *Journal of Politics* 80.1, pp. 240–253.

Schick, Allen (2008). *The Federal Budget: Politics, Policy, Process*. Brookings Institution Press.

Selin, Jennifer L. and Grace Moore (2023). *Keeping Tabs on the Executive*. Forthcoming.

Shipan, Charles R. (2004). "Regulatory Regimes, Agency Actions, and the Conditional Nature of Congressional Influence". *American Political Science Review* 98.3, pp. 467–480.

Silfa, Maria (N.d.). *Legislators and Rulemaking: The Impact of Campaign Contributions*. Working paper, Princeton University; Retrieved 29 April 2023. URL: www.mariasilfa.com/research.

Weingast, Barry R. (1984). "The Congressional-Bureaucratic System: A Principal-Agent Perspective with Applications to the SEC". *Public Choice*, pp. 141–191.

Weingast, Barry R. and Mark J. Moran (1983). "Bureaucratic Discretion or Congressional Control? Regulatory Policymaking by the Federal Trade Commission". *Journal of Political Economy* 91.5, pp. 765–800.

Wilson, Woodrow (1885 [2018]). *Congressional Government*. Outlook Verlag. URL: www.google.com/books/edition/Congressional_Government/N-xRDwAAQBAJ.

Supporting Information for
How Effective is Congressional Oversight of the Bureaucracy?

A Improper Payments Data and Hearings

Table A1: Improper Payments Data Coverage by Agency

Agency	Fiscal Years
Consumer Product Safety Commission	2012–2020
Corporation for National and Community Service (AmeriCorps)	2012–2021
Department of Agriculture	2006–2021
Department of Commerce	2012–2021
Department of Defense	2004–2021
Department of Education	2001–2021
Department of Health and Human Services	2004–2021
Department of Homeland Security	2012–2021
Department of Housing and Urban Development	2003–2021
Department of Interior	2012–2020
Department of Justice	2012–2020
Department of Labor	2002–2021
Department of Transportation	2006–2021
Department of Treasury	2004–2021
Department of Veterans Affairs	2003–2020
Environmental Protection Agency	2012–2020
Federal Communications Commission	2005–2021
General Services Administration	2012–2020
National Aeronautics and Space Administration	2012–2020
National Science Foundation	2012–2020
Office of Personnel Management	2002–2021
Railroad Retirement Board	2003–2020
Small Business Association	2008–2021
Social Security Administration	2000–2021
U.S. Agency for Global Media	2012–2020
U.S. Agency for International Development	2012–2020

B Other tables and figures

In Figure A1 we plot time-series of improper payment rates by program for programs with witnesses called to testify. We re-scale time to years from first oversight hearing for each program. The plot helps us evaluate the assumption of parallel trends. There is little evidence that program trends in IP rates systematically differed in the fiscal years before they were first called to testify. This holds for all programs, as represented by the loess smooth (thick red line), for large programs (with outlays of more than $1 million, thick black lines), and for small programs (thin gray lines).

Figure A1: Improper payment time-series by program. X-axis relative to fiscal year of first oversight hearing.

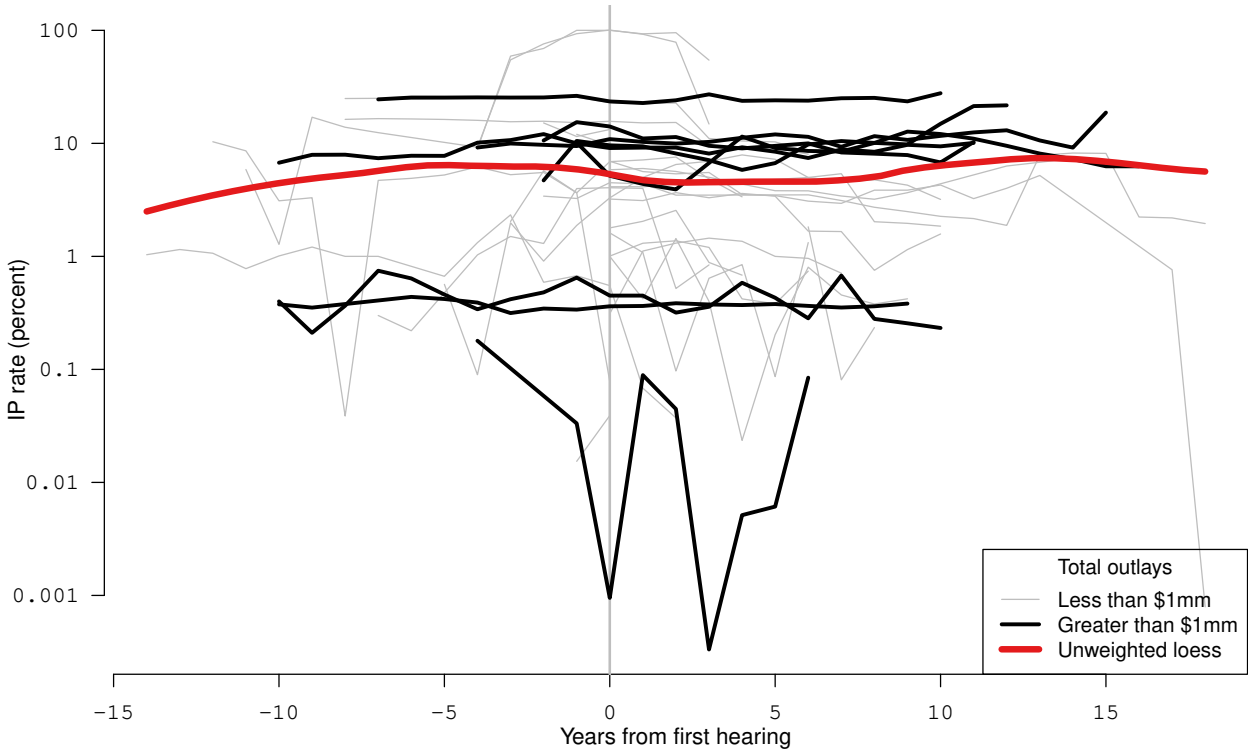


Table A2 provides a regression evaluation of pre-trends. We regress an indicator for oversight hearing in that fiscal year on lags of proportional change in IP rate. Proportional changes in improper payment rates do not predict oversight hearings.

Table A2: Evaluation of Pre-trends. Predicting oversight hearing with prior values of improper payments.

	Agency-level			Program-level		
	(1)	(2)	(3)	(4)	(5)	(6)
Proportional change in IP rate t-1 to t	0.0033			-0.00034		
	(0.0035)			(0.00026)		
Proportional change in IP rate t-2 to t-1		0.00037			-0.00017	
		(0.0035)			(0.00056)	
Proportional change in IP rate t-3 to t-2			0.0071			0.00050
			(0.0090)			(0.00029)
Log improper payment dollars, t-1	0.045	0.042	0.054	0.0049	0.020	0.022
	(0.029)	(0.031)	(0.034)	(0.010)	(0.013)	(0.016)
Fiscal Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Agency/Program FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	257	232	213	939	768	649

Standard errors in parentheses; clustered on agency-year for program-level analysis.

** $p < 0.05$, *** $p < 0.01$

Table A3: Dynamic Effect of Other Moderators on Improper Payment Rate.

	Agency-level					Program-level			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Years Since Last Agency Hearing	-0.028	-0.16***	-1.27	-0.12***	-0.12***	-0.024	-0.20***	-1.48	-0.18***
	(0.16)	(0.041)	(1.13)	(0.033)	(0.035)	(0.21)	(0.065)	(1.18)	(0.053)
Years Since Last * Number Witnesses, Fiscal Year of Last	-0.011					-0.020			
	(0.021)					(0.027)			
Years Since Last * Agency Political Appointee Witnesses, Fiscal Year of Last		0.12**					0.085		
		(0.053)					(0.077)		
Years Since Last * Subcommittee Hearing, Fiscal Year of Last			1.16					1.30	
			(1.13)					(1.16)	
Years Since Last * Pres. Election Year, Fiscal Year of Last				-0.54**	-0.54				-0.58
				(0.27)	(0.28)				(0.49)
Agency Hearing	0.87	0.51	0.12	0.28	0.30	1.81	1.66	1.31	1.09
	(0.66)	(0.52)	(0.92)	(0.51)	(0.54)	(1.09)	(1.06)	(1.01)	(1.01)
Number of Witnesses, This Fiscal Year	-0.052					-0.080			
	(0.090)					(0.077)			
Number of Political Agency Witnesses, This Fiscal Year		0.0060					-0.55		
		(0.40)					(0.58)		
Subcommittee Hearing, This Fiscal Year			0.45					-0.12	
			(0.93)					(0.65)	
Pres. Election year, This Fiscal Year				0.10	0.89				0.035
				(0.99)	(1.28)				(1.86)
Fiscal Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Agency/Program FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	249	249	249	249	234	1150	1150	1150	1150

Standard errors in parentheses; clustered on agency-year for program-level analysis. Column (5) excludes fiscal year 2012, when all improper payments hearings were held during election year 2012.

** $p < 0.05$, *** $p < 0.01$

C Appropriations committee reports

Appropriations committee reports addressing improper payments can cover agencies in two different ways. First, the report can cover an agency and mention that agency's improper payments as needing attention or improvement. For example, the committee report for the 2019 annual appropriations bill for the Departments of Labor, Health and Human Services, and Education singled out the Department of Labor's improper payments:

> The Committee includes $117,000,000 for the RESEA program, and an additional $33,000,000 is made available pursuant to the Bipartisan Budget Act of 2018 (PL 115-123). The Committee continues to support the RESEA program and urges the Department to focus its efforts on reducing the rate of improper payments in the UI system. On June 13, 2017, DOL-OIG issued a report entitled DOL Needs to Do More to Reduce Improper Payments and Improve Reporting (Report Number 03-17-002-13-001). The Committee notes that the rate of improper payments in the UI program exceeds program goals and remains among the highest of any Federal program. The Committee directs ETA to work with DOL-OIG to implement the policies and controls necessary to make significant progress on reducing the rate of improper payments in the UI program.[1]

Or, the report can also cover agencies but not directly refer to an agency's own improper payments specifically. For instance, the committee report for the 2010 Financial Services and General Government Appropriations Bill covered various agencies and mentioned improper payments in the section below:

> The President's budget request proposes a new program, to be managed by the Office of Management and Budget, called the Partnership Fund for Program Integrity Innovation. The purpose of this program is to conduct joint Federal-state pilot projects to demonstrate ways in which the Federal Government and the states can cooperate to improve program administration and services while reducing rates of errors and improper payments, especially for programs in which the states have a substantial role in administration and payments.[2]

This report addressed improper payments in general, and not a specific agency's improper payments.

[1] From committee report number 115-862 accompanying H.R. 6470, issued on July 23, 2018.

[2] From committee report number 111-202 accompanying H.R. 3170, issued on July 10, 2009.

Table A4: Effect of Appropriations Committee Report about Improper Payments on Improper Payment Rate.

	Agency-level				Program-level			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Appropriations report, direct mention	-0.17	-0.21			-1.30**	-1.05		
	(0.53)	(0.49)			(0.59)	(0.62)		
Appropriations report, direct mention, t-1		-0.87				-0.33		
		(0.51)				(0.66)		
Appropriations report			-0.32	-0.31			1.13	1.05
			(0.50)	(0.48)			(1.17)	(1.20)
Appropriations report, t-1				0.23				0.15
				(0.47)				(1.01)
Fiscal Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Agency/Program FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	302	265	302	265	1284	1004	1284	1004

Standard errors in parentheses; clustered on agency-year for program-level analysis.

** $p < 0.05$, *** $p < 0.01$

Figure A2: Improving improper payment rates in Medicare Part C and D. Y-axis is improper payment rate. Points are fiscal year estimated improper payment rates, solid line ordinary least-squares fit.