Lemons in the Political Marketplace: A Big-Data Approach to Detect 'Scam PACs' *

Zhao Li

January 19, 2021

Preliminary and incomplete. Please do not cite or redistribute without permission.

'Scam PACs' are political action committees (PACs) in the United States that raise campaign contributions to enrich their creators (e.g., political consultants) instead of advancing the campaigns or causes they purport to champion. In the 2018 election cycle alone scam PACs collectively raised $57 million, which could have fully funded 74 average House campaigns. The proliferation of and the lack of regulatory oversight over scam PACs not only undermine PACs' accountability to donors, but also generate a lemons problem in the political marketplace. To reduce the information asymmetry that donors face in discerning scam PACs, I first quantitatively assess how scam PACs that have been identified by media reports differ from comparable non-scam PACs on fundraising and expenditure patterns, donor characteristics, and PAC donor and personnel networks. Building on these descriptive analyses, I construct a supervised machine learning algorithm that systematically detects scam PACs in U.S. federal elections.

*Assistant Professor of Politics and Public Affairs at Princeton University. E-mail address: zhaoliresearch@gmail.com. I am thankful for helpful comments from Adam Bonica, David Broockman, Tim LaPira, Greg Martin, Andrew Mayersohn, and participants at the Princeton Research in Experimental Social Science workshop and the annual conference of the American Political Science Association. All errors are my own.

1 Introduction

Principal-agent problems are ubiquitous in life, including in many realms of politics. Indeed, they underlie the most fundamental question in democratic politics of how voters can hold elected officials accountable (e.g., Ferejohn 1986; Persson, Roland, and Tabellini 1997). Looking beyond such canonical cases of principal-agent relationships, political scientists in recent years have increasingly examined the role of political intermediaries in shaping broader outcomes of democratic representation and accountability. For example, legislative staff's ideological orientation and cognitive biases affect their responsiveness to constituents on behalf of Members of Congress (e.g., Furnas 2019; Furnas, LaPira, Hertel-Fernandez, Drutman, and Kosar 2019; Hertel-Fernandez, Mildenberger, and Stokes 2019); profit motives lead lobbyists to alter their efforts to persuade policymakers on behalf of interest group clients (e.g., Drutman 2015; Hirsch, Kang, Montagnes, and You 2020); and campaign consultants' political leanings and material incentives influence the quality and efficiency of their services to client candidate campaigns (e.g., Limbocker and You 2020; Martin and Peskowitz 2015, 2018; Nyhan and Montgomery 2015).

In spite of these recent advances, an important type of principal-agent problems related to intermediaries in politics–the rise of so-called "scam PACs"–has received virtually no attention from political scientists.[1] "Scam PACs" refer to non-connected political action committees (i.e., political action committee, or PACs, with no connections to political candidates, parties, etc.) that solicit campaign contributions from campaign donors with stated goals of supporting specific candidates or political causes, and yet redirect the

[1] A Google Scholar search for "scam PAC" or "scam PACs" returns only 10 results. All of them appear to be qualitative in nature, and none is from political science.

money raised to enrich PAC treasurers, vendors, and other associates.[2] For example, the Tea Party Leadership Fund, an alleged scam PAC, spent roughly 86% of the $6.7 million it has raised since 2013 on consulting firms that assisted the PAC in fundraising, including firms such as DB Capitol Strategies owned by the PAC's treasurer, Dan Backer (Lipton and Steinhauser 2015).

Far from being rare exceptions, scam PACs have proliferated in the post-*Citizens United* era, and increasingly threaten the electoral process in the United States (Raymer 2016; Weintraub and Ravel 2016). For example, in the 2018 federal election cycle alone, PACs that have been alleged to be scam PACs by major news outlets collectively raised more than $57 million in campaign contributions, which could have funded more than 74 average House campaigns in the same cycle.[3] Scam PACs often achieve impressive fundraising success owing to their misleadingly implied associations with popular candidates or causes (e.g., Tea Party Leadership Fund PAC), or their masquerading as charitable organizations (e.g., Children's Leukemia Support Network PAC) (Janetsky 2018; Weintraub and Ravel 2016; Kleiner and Zubak-Skees 2019). At the same time, a substantial degree of information asymmetry has led many donors to fall victim to scam PACs. First, donors are often unaware of publicly available records on PAC expenditures, or do not understand how to take advantage of such informational resources (Hunter, Weintraub, Petersen, and

[2]Formally speaking, a non-connected committee is a political committee that is not a party committee, an authorized committee of a candidate or a separate segregated fund established by a corporation or labor organization (Federal Election Commission 2019a).

[3]The former figure is based on my data collection of scam PACs as detailed in Section 2 as well as my calculation using the FEC's public records. The latter figure is based on the FEC's summary report of the 2018 election cycle, which states that: "[t]he 2,234 candidates running for the House of Representatives reported combined total receipts of $1.7 billion" (Federal Election Commission 2019b).

Walther 2018; Weintraub and Ravel 2016). Second, even if donors familiarize themselves with these resources, they may fail to discern scam PACs due to the lack of a bright line in observable conduct that separates scam PACs from legitimate PACs (Janetsky 2018; Kleiner and Zubak-Skees 2019). For example, treasurers of scam PACs that divert most of the contributions they raise to fundraising often defend their conduct as a necessity of building the groundwork for newly founded PACs (even if doing so conveniently creates opportunities for financial self-dealing) (Lipton and Steinhauser 2015; Severns and Willis 2019). Third, the Federal Election Commission, the primary regulatory agency over federal campaign finance activities, has little authority in reigning in scam PACs under existing campaign finance laws (Hunter, Weintraub, Petersen, and Walther 2018), and recent court cases have further deprived the Commission of its ability to combat even egregious cases of scam PACs that misrepresent themselves as official candidate campaigns (e.g., *Pursuing America's Greatness v. FEC*).

Political scientists can help to ameliorate this lemons problem in the political marketplace by providing informational tools for donors to detect scam PACs. Such efforts would not only address an important need for election administration (Hunter, Weintraub, Petersen, and Walther 2018; Weintraub and Ravel 2016), but also shed light on an under-explored type of principal-agent problems that relate to PACs' accountability to campaign donors. In the context of campaign finance, existing research largely examines principal-agent problems in terms of politicians' accountability to voters (i.e., can campaign contributions corrupt elect officials and erode the representation of constituent interests?) (e.g., Bartels 2012; Lessig 2011), or politicians' accountability to donors (i.e., can donors get what they want by making campaign contributions to candidates?) (e.g., Ansolabehere, de Figueiredo, and Snyder 2003; Kalla and Broockman 2016). In contrast, with few exceptions related to corporate governance (e.g., Li 2018; Min and You 2019), the extent to which PACs as political intermediaries are accountable to their donors remains an open question. At the same time, this question is integral to our understanding of the contemporary campaign fundraising landscape, especially as outside spending continues to thrive in the post-*Citizens United v. FEC* era.[4]

In the case of scam PACs, campaign donors' inability to discipline scam PACs, either via direct intervention or indirectly by "voting with their money", not only undermines donors' ability to achieve their political goals through campaign contributions (which scam PACs siphon off from the candidates or causes that donors support), but also generates broad-ranging negative externalities in political fundraising. As awareness of the problem of scam PACs spreads across donors, the challenges of differentiating scam PACs from candidate campaigns or legitimate PACs could lead donors to become disillusioned and withdraw from making campaign contributions altogether (Severns and Willis 2019), which would further undercut fundraising for candidate campaigns and legitimate PACs by shrinking the donor pool (beyond losses in campaign contributions that they already incur due to competition from scam PACs). Moreover, election administrators fear that inexperienced donors could be especially likely to exit in the presence of scam PACs (Weintraub and Ravel 2016), which could threaten to undo recent progress in the diversification of the donor pool and further exacerbate inequality in participation in campaign finance (e.g., Grumbach and Sahn 2020; Grumbach, Sahn, and Staszak N.d.).

To reduce the type of information asymmetry in political fundraising that enables scam PACs to proliferate, my paper provides a first attempt at helping campaign donors discern scam PACs. To this end, I start with descriptive analyses that compare scam PACs to non-scam PACs on a variety of observable attributes. Section 2 details how I construct my data sample for both types of PACs, and Section 3 shows that scam PACs differ from non-scam PACs in several aspects, including fundraising and expenditure patterns (e.g., fundraising size, itemization ratio, budget allocation across expense categories), donor characteristics (e.g., ideology and age), and the sets of donors, PAC treasurers, and ven-

[4]https://www.opensecrets.org/outsidespending/cycle_tots.php

dors to which each PAC is connected. Building upon these descriptive findings, I then construct a supervised algorithm that systematically detects scam PACs in federal elections based on publicly available campaign finance data, using a set of scam PACs that have been reported by major news outlets as the training set for model estimation. As shown in Section 4, preliminary results are promising, and suggest that supervised machine learning has the potential to help donors distinguish scam PACs (including those that are new or have yet to be flagged in government or media reports) from non-scam PACs without relying on arbitrary and often subjective rules of thumbs in classifying scam PACs (Janetsky 2018; Kleiner and Zubak-Skees 2019). Last but not least, Section 5 concludes by highlighting potential areas of improvement in the detection of scam PACs, as well as future research that could build upon this paper in assessing how equipping donors with such informational tools may change donor behavior and ameliorate the lemons problem in the political marketplace.

2 Data Construction

2.1 Data sources and time frame

I collect publicly available federal campaign finance records from the following three sources: the Federal Election Commission's bulk data depository, the Center for Responsive Politics (CRP)'s bulk data depository, and the Database on Ideology, Money in Politics, and Elections (DIME) (Bonica 2019). For now, I focus on records compiled for the 2010 through 2018 federal election cycles, since all known scam PACs were founded in the past decade, and that the 2020 data are not yet complete. However, the time range of my data collection certainly can be extended backwards if I uncover more likely cases of scam PACs that occurred earlier, and forwards as new scam PACs emerge in the current decade.

2.2 *Identifying scam PACs*

To collect a sample of scam PACs, I read through more than 30 reports or documents by major news outlets (e.g., *New York Times*, *Wall Street Journal*, *Politico*) and campaign finance-focused non-profit organizations (e.g., the Center for Public Integrity, OpenSecrets) that allege specific PACs as scam PACs and provide substantiating evidence. Table 1 lists the 46 alleged scam PACs that I have identified through this procedure, along with hyper-linked sources. This list will expand the sample as I parse through additional sources.

For each of these 46 scam PACs, I identify their FEC committee ID's by PAC name, and then compile data on each scam PAC's fundraising and expenditures, data on their itemized campaign donors, as well as data on their PAC treasurers and vendors from the FEC, CRP, and DIME. Using the data thus collected, I then identified $51,147$ unique campaign donors (using *bonica.cid* in DIME as time-invariant donor identifiers) who have made at least one itemized contribution to one or more scam PACs within the time frame of my analysis.

Table 1: A Non-Exhaustive List of Alleged Scam PACs

Index	PAC Name	Source(s)
1	American Coalition for Injured Veterans	Center for Public Integrity
2	Americans for Law Enforcement	Campaigns and Elections
3	Americans for Police and Trooper Safety	Politico
4	Americans for the Cure of Breast Cancer	Politico, Center for Public Integrity
5	Americans Socially United	OpenSecrets
6	Association for Emergency Responders and Firefighters PAC, Inc.	Politico, Center for Public Integrity
7	Autism Hear Us Now PAC	Center for Public Integrity
8	Bold Conservatives PAC	OpenSecrets
9	Coalition of Americans for Political Equality	Sunlight Foundation
10	Committee to Restore America's Greatness	OpenSecrets
11	Community Health Council	Center for Public Integrity
12	Conservative Action Fund	New York Times, Wall Street Journal
13	Conservative America Now	Wall Street Journal
14	Conservative Freedom Fighters	OpenSecrets
15	Conservative Majority Fund	OpenSecrets, Politico
16	Conservative Strikeforce	Wall Street Journal, OpenSecrets
17	Constitutional Rights PAC	New York Times
18	Cops and Kids Together	Politico
19	Children's Leukemia Support Network	Center for Public Integrity
20	Firefighters' Alliance of America	Center for Public Integrity
21	Freedom's Defense Fund	OpenSecrets
22	Grassroots Awareness PAC	Campaigns and Elections
23	Great America PAC	Politico
24	Heart Disease Network of America	Center for Public Integrity
25	Heroes United	Center for Public Integrity
26	Law Enforcement for a Safer America	WUSA (CBS)
27	Life and Liberty PAC	Roll Call
28	Madison Project	New York Times
29	National Assistance Committee	Politico
30	National Campaign PAC	Campaigns and Elections
31	Patriot Super PAC	Sunlight Foundation
32	Police Officers Defense Alliance	Center for Public Integrity
33	Put Vets First!	Center for Public Integrity
34	Remember Mississippi	CBS
35	Republican Majority Campaign PAC	OpenSecrets
36	Restore American Freedom and Liberty	Politico
37	Standing by Veterans PAC, Inc.	Politico, Center for Public Integrity
38	Tea Party Leadership Fund	New York Times
39	Tea Party Majority Fund	OpenSecrets
40	Tea Party Patriots	New York Times
41	Tea Party Victory Fund	OpenSecrets
42	United Police Officers Association	Center for Public Integrity
43	United Veterans Alliance of America PAC Inc	Center for Public Integrity
44	US Veterans Assistance Foundation	Politico, Center for Public Integrity
45	Virgin Islands GOP	OpenSecrets
46	Volunteer Firefighters and Paramedic Association	Center for Public Integrity

2.3 *Identifying comparable non-scam PACs*

Next, in order to construct valid comparison groups for these scam PACs, I create a list of what I refer to as "non-scam PACs", i.e., PACs that have not been reported to engage in fraudulent fundraising conduct and otherwise fall into similar categories as most scam PACs. Specifically, each non-scam PAC in my sample must satisfy the following criteria: 1) it must not be authorized by any candidate campaigns; 2) it must not be a party committee; and 3) it must not be a segregated separate fund (i.e., sponsored by a corporation or a union) (Federal Election Commission 2017). These criteria ensure that the non-scam PACs in my sample have comparable organizational structures, donor bases, and fundraising and expenditure needs as scam PACs (at least more comparable than other PACs that do not meet these criteria). In total, between the 2010 and 2018 election cycles, there were $3,779$ non-scam PACs per my definition, which received itemized campaign contributions from $1,020,547$ donors.

Since scam PACs that have a clear partisan bias overwhelmingly claim to align with Republican candidates or conservative movements Lipton and Steinhauser (2015), and that donor attributes differ significantly by partisanship and ideology (Barber, Canes-Wrone, and Thrower 2017), I further define a subset of non-scam PACs that are conservative-leaning (which I thereafter refer to as "conservative non-scam PACs") if the itemized donors of these PACs have an average or median contributor CFscore above zero (Bonica 2014). Per my definition, from the 2010 federal election cycle to the 2018 federal election cycle, there were 932 conservative non-scam PACs, which received itemized campaign contributions from $238,191$ donors.

3 Descriptive Analyses

While the ultimate goal of this paper is to construct a supervised machine learning algorithm that can detect scam PACs, I first conduct a series of descriptive comparisons of scam PACs vs. non-scam PACs on a set of salient characteristics. These descriptive comparisons can help to inform my machine learning exercise (specifically my choices of feature selection and extraction) by highlighting the set of PAC attributes that may best distinguish scam PACs from non-scam PACs. This section reports these descriptive findings.

3.1 PAC fundraising and expenditures

For each of the three types of PACs that I examine–scam PACs, (all) non-scam PACs, and conservative non-scam PACs–I calculate a set of metrics that relate to PAC fundraising and expenditures. I measure for each PAC its average total amounts of money raised vs. spent per active cycle. In addition, informed by FEC reports highlighting the markedly lower itemization ratios of campaign contributions raised by scam PACs relative to those of non-scam PACs (Weintraub and Ravel 2016), I compute for each PAC its average itemization ratio in fundraising per active cycle. Last but not least, since a large number of government as well as media reports highlight the distinct expenditure patterns of scam PACs (Graham 2019; Janetsky 2018; Kleiner and Zubak-Skees 2019; Weintraub and Ravel 2016), I also calculate for each PAC its average percentage of expenditures per active cycle on various expense categories as coded by the CRP.[5]

[5] The FEC also codes expenditures according to its own classification system. However, a substantial portion of expenditures remains unclassified in FEC records. Analysis based on FEC categories generates qualitatively comparable results.

Table 2: PAC-Level Summary Statistics

Statistic	Scam PACs	Non-Scam PACs	Conservative Non-Scam PACs
No. PACs	46	3,779	932
Ave. total fundraising per cycle	$1,760,253	$416,803	$895,590
Ave. total expenditure per cycle	$1,972,108	$861,408	$1,109,823
Ave. itemization ratio per cycle	0.266	0.775	0.813
% Expenditure on Contributions (CRP)	4.3%	16.4%	15.2%
% Expenditure on Unclassifiable (CRP)	11.5%	9.4%	8.4%
% Expenditure on Administrative (CRP)	15%	19.7%	20.9%
% Expenditure on Non-Expenditures (CRP)	2.4%	3.8%	4.3%
% Expenditure on Transfers (CRP)	0.1%	8.5%	6.3%
% Expenditure on Strategy & Research (CRP)	0.018	0.07	0.073
% Expenditure on Campaign Expenses (CRP)	7.9%	3.6%	2.8%
% Expenditure on Salaries (CRP)	1.6%	2.4%	1.9%
% Expenditure on Fundraising (CRP)	43.3%	11.4%	14.1%
% Expenditure on Media (CRP)	5.5%	8.6%	8.3%

Three sets of findings emerge from my descriptive analysis of PAC fundraising and expenditure patterns as reported in Table 2. All of the comparisons here hold even if we restrict attention to conservative non-scam PACs as the comparison set, which suggests that partisan or ideological leanings alone are unlikely to account for the distinct fundraising and expenditure patterns of scam PACs.

First, scam PACs had much larger budgets overall (almost $2 million per active cycle), as measured by both their average total fundraising and average total expenditure. Non-scam PACs typically operated on budgets that were less than half the size. This may reflect the fact that the set of scam PACs that has attracted major media attention tends to be more successful at fundraising.

Second, only about 26.6% of all campaign contributions raised by scam PACs in a typical cycle were itemized, which is less than a third of that for comparable non-scam PACs.

Third, scam PACs appeared to pursue a markedly different strategy for campaign expenditures compared to non-scam PACs. Specifically, on average almost half (43.3%) of scam PACs' expenditures went to fundraising, which is, proportionally speaking, roughly four times as much as that of non-scam PACs. At the same time, scam PACs spent little on making campaign contributions (4.3%) or transferring their campaign funds (0.1%), which collectively constituted more than a fifth of most non-scam PACs' expenditures in a given active cycle. Last but not least, though the difference here is less pronounced, scam PACs also appeared to engage in less electioneering activities (e.g., media, strategy and research).

3.2 *Itemized Donors*

In addition to examining patterns of PAC fundraising and expenditures, I investigate traits of itemized donors of each type of PACs: scam PACs, (all) non-scam PACs, and conservative scam PACs. Journalistic accounts highlight that elderly donors appear to be more likely to fall victim to scam PACs (Graham 2019). While itemized donors are not required to report their age to the Federal Election Commission, they are asked about their occupations (though disclosure is self-reported and not verified). To proxy for age, I calculate the fraction of itemized donors that claim to be retirees for each type of PACs examined. Moreover, since media reports point to the notable conservative orientation of many scam PACs (Severns and Willis 2019), I examine the average contributor CFscores (Bonica 2014) of itemized donors for each type of PACs, where higher values of contributor CFscores correspond to greater conservative leaning in a donor. Furthermore, some media reports suggest that donors who give to scam PACs may be less habitual donors in general (Arnsdorf and Vogel 2016; Severns and Willis 2019). To measure donors' experience with campaign contributions, I calculate the average total number of distinct recipients (across different categories) as well as the average total number of active cycles of giving for itemized donors of scam PACs, (all) non-scam PACs, and conservative scam PACs, respectively.

Table 3: PAC Donor-Level Summary Statistics

Statistic	Donors to Scam PACs	Donors to Non-Scam PACs	Donors to Conservative Non-Scam PACs
No. unique donors	54,147	1,020,547	238,191
% retirees	53%	9.1%	39.2%
Ave. donor CFscore	1.31	-1.303	1.279
Ave. no. scam PACs given to	1.117	0.022	0.094
Ave. no. non-scam PACs given to	0.794	1.327	1.311
Ave. no. conservative non-scam PACs given to	0.786	0.303	1.296
Ave. no. all recipients given to	9.711	12.649	10.937
Ave. no. active cycles	2.493	2.47	2.468

Table 3 report compares these key measures of donor behavior for itemized donors of scam PACs, (all) non-scam PACs, and conservative scam PACs, respectively. Since publicly available campaign finance records limits my analysis to only *itemized* donors of scam PACs, i.e., those who have donated $200 or more to at least one scam PAC in an election cycle, conclusions drawn from Table 3 need not generalize to all scam PAC donors (the vast majority of whom, as shown in Table 2, donate much less than the itemization threshold).

First, itemized donors of scam PACs appear much older. 53.0% of these donors were self-reported retirees according to their itemized campaign contribution records. This contrasts with 39.2% of self-reported retirees among itemized donors to conservative non-scam PACs, and 9.1% for all non-scam PACs.

Second, itemized donors of scam PACs tend to be much more conservative. The average contributor CFscore for these donors is 1.310, which is higher (i.e., more conservative according to DIME's scaling) than that of the average itemized donor to conservative non-scam PACs (1.279), and especially so compared to the average itemized donor across all non-scam PACs (-1.303).

Third, itemized donors who gave to scam PACs may be at best slightly less habitual in campaign giving compared to itemized donors of comparable non-scam PACs. Itemized donors of scam PACs did not exclusively contribute to scam PACs; they donated on average to 0.794 non-scam PACs, and 0.786 conservative non-scam PACs (although these numbers are roughly a half of those of itemized donors of non-scam PACs). Moreover, in terms of the number of all recipients (candidates, PACs, etc.) to whom a given itemized donor has contributed, itemized donors of scam PACs are nearly identical to donors of conservative non-scam PACs (on average around 10), and only slightly behind donors of all non-scam PACs (approximately 12 on average). Finally, itemized donors of scam PACs made itemized contributions (to any recipient) across roughly 2.49 election cycles, which is almost identical to that of itemized donors of comparable non-scam PACs.

3.3 PAC treasurers and vendors

In the last set of descriptive analyses, I turn my attention to the personnel behind PACs' operations, specifically their treasurers and vendors. Journalistic accounts of scam PACs suggest that their treasurers and vendors are often veterans in political consulting, and leverage their expertise as well as connections to found or serve scam PACs as a means of financial self-dealing (Arnsdorf and Vogel 2016; Janetsky 2018; Kleiner 2017; Kleiner and Zubak-Skees 2019; Lipton and Steinhauser 2015; Severns and Willis 2019). If these accounts are representative of scam PACs, we should expect treasurers and vendors of scam PACs to serve a greater number of PACs compared to their peers that only serve non-scam PACs.

To test this idea, I first collect data on names and addresses of PAC treasurers as reported to the FEC (none of the existing data sources on campaign finance records–FEC, CRP, or DIME–provides identifiers for unique PAC treasurers). I then standardize these names, using a set of string cleaning procedures customized for this data set, so that these standardized names may serve as identifiers of PAC treasurers, and disambiguiating identities using address information where appropriate. I examine three sets of PAC treasurers: those that served as treasurers of scam PACs, (all) non-scam PACs, and conservative scam PACs, respectively (these categories are far from mutually exclusive). For each of these sets of PAC treasurers, I compute the number of scam PACs, (all) non-scam PACs, and conservative scam PACs that they have served as treasurers for. Tables 4 reports these numbers.

Table 4: PAC Treasurer-Level Summary Statistics

Statistic	Treasurers of Scam PACs	Treasurers of Non-Scam PACs	Treasurers of Conservative Non-Scam PACs
No. unique treasurers	22	1,677	617
Ave. no. scam PACs served	1.545	0.014	0.037
Ave. no. non-scam PACs served	3.227	1.149	1.305
Ave. no. conservative non-scam PACs served	3.091	0.466	1.267

As shown in Table 4, I identified 22 unique scam PAC treasurers. Note that, among them, Scott B. MacKenzie, Dan Backer, Alexander Hornaday, and Paul Kilgore have sponsored multiple scam PACs, which correspond well to existing journalistic accounts of these individuals' high-profile involvement in scam PACs (Arnsdorf and Vogel 2016; Janetsky 2018; Lipton and Steinhauser 2015). These 22 scam PAC treasurers have performed the same role for, on average, 3.227 non-scam PACs, including 3.091 conservative non-scam PACs. These numbers are about twice to three times as large as those for treasurers of non-scam PACs, which demonstrates that scam PAC treasurers indeed appeared to have more experience in campaign fundraising.

I also calculate an analogous set of summary statistics for PAC vendors. Similar to the case of PAC treasurers, none of the existing campaign finance data sources includes identifiers of unique PAC vendors. While I have not finished standardizing PAC vendor names, I have done so for the top 50 vendors of scam PACs and non-scam PACs, respectively, and Table 5 reports preliminary results based on my current stage of progress.

Table 5: PAC Vendor-Level Summary Statistics

Statistic	Vendors of Scam PACs	Vendors of Non-Scam PACs	Vendors of Conservative Non-Scam PACs
No. unique vendors	1,452	23,945	15,113
Ave. no. scam PACs served	1.433	0.047	0.074
Ave. no. non-scam PACs served	5.269	1.625	1.902
Ave. no. conservative non-scam PACs served	3.749	1.023	1.621

As shown in Table 5, I identified $1,452$ unique vendors that have served one or more scam PACs in my sample. Similar to the case of PAC treasurers, here scam PAC vendors also served more clients in general: on average 5.269 non-scam PACs, including 3.749 conservative non-scam PACs. These numbers are two to three times as large as those for vendors that served non-scam PACs. In addition, the identities of scam PAC vendors also corroborate my analysis of scam PACs' expenditure patterns as discussed in Section 3.1. For example, the top five most frequent vendors among scam PACs are the U.S. Postal Service, American Technology Services, United Data Services, Compliance Consultants, and Paypal. These observations are consistent with results reported earlier in Table 2, revealing that scam PACs allocate a much greater fraction of their total expenditures to fundraising (including expenses for maintaining contact lists and mailing as well as digital marketing campaigns) and other expenses that do not directly affect campaigns or elections.

4 Supervised Machine Learning

The key features detailed in the previous section that distinguish scam PACs from non-scam PACs help to inform my construction of a supervised machine learning algorithm that systematically detects likely scam PACs. Such an algorithm would directly address the need for an independent source of information that helps potential donors to discern scam PACs from non-scam PACs based on publicly available records of PACs, as urged by the Federal Election Commission (Weintraub and Ravel 2016). In particular, insofar as high out-of-sample accuracy is obtainable, machine learning-based predictions of PACs' likelihood to be scam PACs can help donors identify scam PACs that are new or have yet to receive public scrutiny, thereby expanding the set of detectable scam PACs beyond those that have already been flagged by investigative journalists. In short, a supervised algorithm for scam PAC detection has the potential of ameliorating significant information asymmetry in the fundraising marketplace, which is one of the key causes for the proliferation of scam PACs (Hunter, Weintraub, Petersen, and Walther 2018; Weintraub and Ravel 2016). The rest of this section describes how I estimate such supervised models, as well as analyzing the predictions I obtain from these models.

4.1 Sample selection

I start with the set of 46 scam PACs and $3,779$ non-scam PACs that I identified using the procedure outlined in Section 2. Since the typical operating budgets of many non-scam PACs were much smaller than those for scam PACs, as shown in Table 2, I first drop PACs whose average total expenditures per active cycle were less than $\$10,000$. This leaves me with 45 scam PACs and $1,439$ non-scam PACs. Later, after constructing the set of model features (i.e., predictors) for supervised machine learning, I drop an additional scam PAC as well as 442 additional non-scam PACs due to missing data present in model features. This leaves me with 44 scam PACs and 997 non-scam PACs.

4.2 Target outcome

The outcome variable that I set out to predict using supervised machine learning is whether a given PAC is a scam PAC versus a non-scam PAC, as well as the probabilities that a given PAC falls into each of the two categories. For now I label all PACs in my data set for model training as either scam PACs (i.e., having been flagged by government or media reports) or non-scam PACs.

4.3 Supervised algorithm

The objective of this machine learning application is to identify scam PACs based on publicly available records about their fundraising and expenditure patterns, their donors, their treasurers, and their vendors. Even though the descriptive findings described in Section 3 highlights a number of observable aspects in which scam PACs differ from non-scam PACs, aggregate statistics alone may not be sufficient for predicting scam PACs with a high degree of accuracy. More importantly, in the absence of a clear legal definition for scam PACs, existing methodologies for identifying scam PACs, found in investigative reports by journalists, largely rely on arbitrary rules of thumbs that both lack validation and invite skepticism on the grounds of subjectivity (Janetsky 2018; Kleiner and Zubak-Skees 2019). To systematically detect scam PACs, supervised machine learning is a suitable methodological approach as it is designed to "learn" unobserved and likely complicated mappings from predictors to target outcomes.

To formally describe my supervised algorithm, let N_{train} be the set of PACs whose type (i.e., scam vs. non-scam PACs), $\mathbf{Y}_{train}$, are defined. Then, let $\mathbf{W}_{train}$ be an $N_{train} \times m$ matrix whose m columns represent model predictors (to be described in Section 4.4). Let $f(\cdot)$ be the unobserved function that best summarizes how model predictors map onto PAC types for the training set:

$$\mathbf{Y}_{train} = f(\mathbf{W}_{train}) \tag{1}$$

Supervised machine learning estimates a function $\hat{f}(\cdot)$ that best approximates the true mapping $f(\cdot)$. With $\hat{f}(\cdot)$, I can then use it to predict whether any PAC not included in the training set is more likely to be a scam PAC or a non-scam PAC, as well as the probabilities that it is in each of the two categories. Let $N_{i \in test}$ denote the set of PACs in the test set (i.e., held out from model estimation). Their respective predictive PAC types are thus

$$\hat{\mathbf{Y}}_{test} = \hat{f}(\mathbf{W}_{test}) \tag{2}$$

To estimate $\hat{f}(\cdot)$, I use the *caret* package in *R* (Kuhn 2008) to implement a random forest model for each issue. As a type of decision-tree based algorithms, random forest models are resistant to over-fitting (Breiman 2001), which is important in this application: insofar as scam PACs that I have not included in my data set may differ in systematic ways, an over-fitted supervised algorithm would have limited predictive power for out-of-sample cases, which would defeat the purpose of using supervised machine learning to systematically detect scam PACs whether or not they have received media coverage. In addition, random forest models have built-in estimates of variable importance, which helps to identify specific model predictors that provide the most marginal information on whether a given PAC is a scam or non-scam PAC.

4.4 Feature selection

I use a variety of model features to predict whether a given PAC is a known scam PAC or not. First, I include a set of PAC-level covariates, based on descriptive findings shown in Table 2. For each PAC, these covariates are: the election cycles between 2010 and 2018 in which the PAC was active, the average amounts of total fundraising and expenditures in a given active cycle, the average itemization ratio of fundraising in a given active cycle, and percentages of expenditures allocated to different categories based on both the FEC's and the CRP's classification systems.

Second, I include three aggregate measures of itemized donors that have contributed to each PACs, following the conclusions drawn from Table 3: the share of itemized donors of a given PAC that self-reported as retirees, and the mean and median contributor CFscores associated with itemized donors of each PAC.

Third, I construct a matrix of donor-PAC ties, in which each row is a given PAC, each column is an itemized donor, and each cell–which takes either value of $\{0, 1\}$–indicates whether a given donor has given one or more itemized donations to a PAC. Such donor-recipient matrices, when applied to studies of legislative behavior, have helped to produce highly accurate predictions of federal candidates' DW-NOMINATE scores as well as issue-specific positions (Bonica 2018; Bonica and Li 2019). Donor-recipient matrices enhanced predictive power in these existing applications since donors are discerning of candidates' ideologies and policy platforms (Barber, Canes-Wrone, and Thrower 2017). In the case of this paper, while my outcome variable of interest is not ideology-based, many scam PACs do attempt to appeal to conservative donors (Lipton and Steinhauser 2015), which is corroborated by Table 3. Moreover, other individual donor characteristics (e.g., age), including traits that may not be observable to the researcher, could affect donors' propensities to donate to scam PACs. A key advantage of including such a donor-PAC matrix as I described is that as long as certain donors are more likely to contribute to scam PACs for any reason, donor-PAC linkages based on itemized contribution records can help to detect scam PACs in a supervised machine learning framework.

Last but not least, analogous to the donor-PAC matrix just described, I include a matrix of donor-treasurer ties and another matrix of donor-vendor ties. Since certain PAC treasurers and vendors are more involved in scam PACs than others (Arnsdorf and Vogel 2016; Janetsky 2018; Kleiner 2017; Kleiner and Zubak-Skees 2019; Lipton and Steinhauser 2015; Severns and Willis 2019), an observation supported by Tables 4 and 5, we should expect PACs' links to individual treasurers and vendors to also enhance the predictive performance of my supervised algorithm.

Since the complete donor-PAC, treasurer-PAC, and vendor-PAC matrices are highly sparse (i.e., the typical donor/treasurer/vendor is associated with very few PACs), I drop donors, treasurers, or vendors that are linked to fewer than 8 PACs (including both scam and non-scam PACs) in the training set, which corresponds to roughly 1% of all training observations. While doing so reduces the number of model features, it considerably lowers the computational cost of model estimation. This leaves me with 2, 361 unique donors, 5 treasurers, and 292 vendors.

4.5 Model Fitting

I randomly selected 3/4 of the scam PACs as well as 3/4 of the non-scam PACs to be included in my training data. The rest was held out as the test set.

To train my random forest model, I use repeated 10-fold cross-validation. In other words, the estimation procedure partitions the training set into 10 groups and repeatedly fits the model each time while holding one of the 10-sets out of sample.

For the purpose of model tuning, these repeated cross-validation runs help me to choose the optimal value on the number of variables to be randomly sampled at each split during random forest estimation. The optimal value turns out to be 243, which corresponds to about 9% of the total number of predictors.

4.6 Estimation Results

In this subsection, I describe results from my estimated random forest model. First, I evaluate out-of-sample model performance by assessing predictions for the held-out test data set. Since my data sample is heavily unbalanced (i.e., most PACs are classified as non-scam PACs), a natural concern is that an algorithm that predicts every PAC to be a non-scam PAC can achieve a high degree of predictive accuracy without detecting any scam PACs. To alleviate this concern, I assess model performance based on the area under curve (AUC) metric, which "does not have any bias toward models that perform well on the majority class at the expense of the majority class" (He and Ma 2013, p. 27). In addition, the corresponding confusion matrix displays frequencies of both false positives and false negatives.

Figure 1 displays the receiver-operating-characteristic (ROC) curve for out-of-sample predictions of scam PACs. The AUC in this case is 0.891. As a benchmark, if one naively predicts all PACs to be non-scam PACs, the implied AUC would be 0.5 (i.e., no discrimination across different PAC types). In most machine learning applications to classification problems, an AUC of 0.90 or above is considered "excellent" (Hosmer, Lemeshow, and Sturvidant 2013, p. 177), which is just above the level achieved by my random forest model.

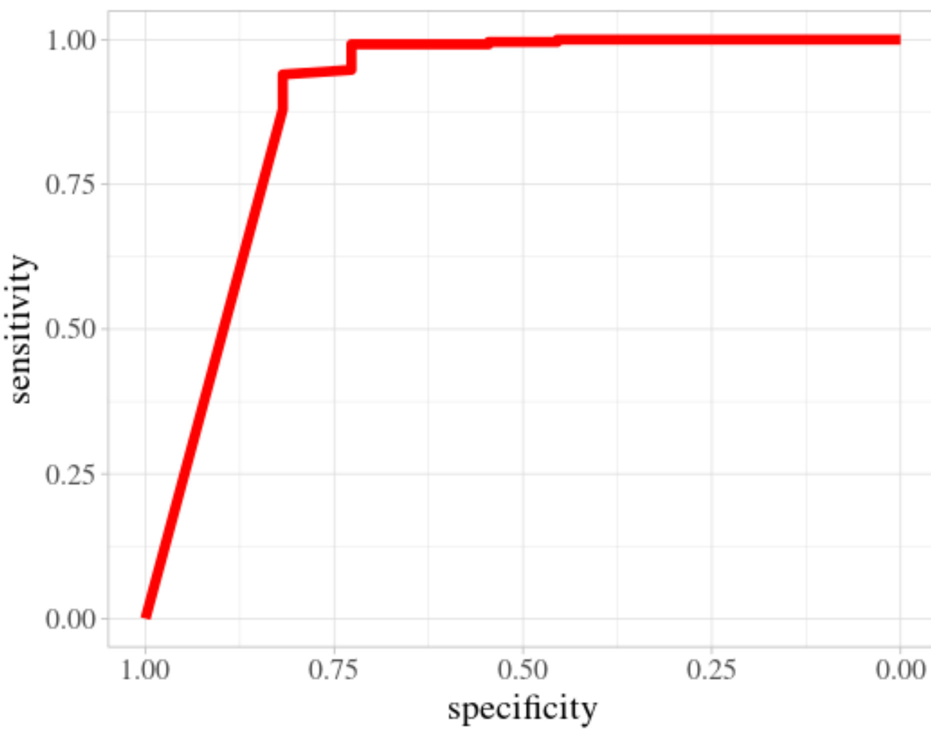

Figure 1: ROC Curve for Out-of-Sample Predictions of Scam PACs

In addition, Table 6 shows the confusion matrix for out-of-sample predictions based on my model. As shown in this table, the model does not naively predict all PACs in the test data set to be non-scam PACs. Instead, it is able to distinguish scam PACs from non-scam PACs, albeit with errors. The false positive rate here is 22.2% (i.e., 2 out of 9), and the false negative rate is 1.6% (i.e., 4 out of 251). This false positive rate may seem high at first glance. However, since I classify all PACs as non-scam PACs by default unless they have been mentioned in one of the news or government reports on scam PACs that I have collected, some of the false positive cases may simply be *bona fide* scam PACs that either escaped my ongoing and incomplete data collection efforts, or are too new to be reported by major media outlets. In particular, when I examine the 2 non-scam PACs in the test data set that are mis-classified as scam PACs by my random forest model, one of them–"For A Better America" PAC–turns out to be an alleged scam PAC that I have yet to add to my data sample (Renshaw and Tanfani 2020), which speaks to the potential for systematic detection of scam PACs based on my supervised algorithm.[6] The full list of PACs predicted to be scam PACs by my random forest model, combining PACs from both the training and test data sets, are shown in Table 7 in descending order of predicted probabilities.

Table 6: Out-of-Sample Confusion Matrix

Prediction \ Reference	legit	scam
legit	247	4
scam	2	7

[6]The other false positive in the test data set is "FreedomWorks for America" PAC, which by all available accounts appears to be a true false positive.

Table 7: List of All PACs Predicted As Scam PACs

PAC Name	Sample	Observed PAC Type	Predicted PAC Type	Predicted Probability of Being a Scam PAC
Cops and Kids Together	training data	scam	scam	1
Autism Hear Us Now PAC	training data	scam	scam	1
Americans for Police and Trooper Safety	training data	scam	scam	1
Life and Liberty PAC	training data	scam	scam	0.952
Conservative America Now	training data	scam	scam	0.905
Tea Party Majority Fund	training data	scam	scam	0.905
Great America PAC	training data	scam	scam	0.905
Us Veterans Assistance Foundation	training data	scam	scam	0.905
Firefighters' Alliance Of America	training data	scam	scam	0.905
National Assistance Committee	training data	scam	scam	0.857
Standing By Veterans PAC, Inc.	training data	scam	scam	0.857
Conservative Majority Fund	training data	scam	scam	0.857
Heroes United	training data	scam	scam	0.857
Americans for The Cure Of Breast Cancer	test data	scam	scam	0.857
United Veterans Alliance Of America PAC Inc	test data	scam	scam	0.857
National Campaign PAC	training data	scam	scam	0.81
Americans for Law Enforcement	training data	scam	scam	0.81
Freedom's Defense Fund	training data	scam	scam	0.81
United Police Officers Association	training data	scam	scam	0.81
Association for Emergency Responders and Firefighters PAC, Inc.	test data	scam	scam	0.81
Put Vets First!	training data	scam	scam	0.762
Republican Majority Campaign PAC	training data	scam	scam	0.762
Tea Party Leadership Fund	training data	scam	scam	0.762
Patriot Super PAC	training data	scam	scam	0.762
Grassroots Awareness PAC	training data	scam	scam	0.762
Conservative Action Fund	training data	scam	scam	0.762
Heart Disease Network Of America	test data	scam	scam	0.762
Community Health Council	training data	scam	scam	0.714
Conservative Freedom Fighters	training data	scam	scam	0.714
Children's Leukemia Support Network	test data	scam	scam	0.714
Restore American Freedom and Liberty	training data	scam	scam	0.667
American Coalition for Injured Veterans	training data	scam	scam	0.667
Remember Mississippi	training data	scam	scam	0.667
Bold Conservatives PAC	training data	scam	scam	0.667
Virgin Islands GOP	training data	scam	scam	0.667
Tea Party Victory Fund	training data	scam	scam	0.667
Freedomworks for America	test data	non-scam	scam	0.667
Volunteer Firefighters and Paramedic Association	training data	scam	scam	0.619
Law Enforcement for A Safer America	test data	scam	scam	0.619
Conservative Strikeforce	training data	scam	scam	0.571
for A Better America	test data	non-scam	scam	0.571
Police Officers Defense Alliance	test data	scam	scam	0.524

Using a built-in algorithm for the random forest model, I also assess which model features are the most "important" variables in the estimated model i.e., they provide the greatest marginal improvement in predictive accuracy. In classification problems, variable importance relates to node impurity (analogous to residual sum of squares in regressions), which is often measured by Gini coefficients. Each model feature's mean decrease in Gini coefficients averages the reduction in Gini coefficients across all nodes where said variable is used for node splitting, and thus intuitively captures the degree of *unique* information that a variable adds to the algorithm. Model features with relatively low mean reduction in Gini coefficients may either be uninformative in distinguishing scam PACs from non-scam PACs, or that the information they provide is duplicated by that of other model features.

Table 8 displays the list of top 50 model features in terms of variable importance, measured by mean decrease in Gini coefficients. In addition to summary statistics of PACs and PAC donors that appear to distinguish scam PACs from non-scam PACs, consistent with results shown in Section 3, this table also suggests that PACs' linkages to individual donors and vendors help to differentiate PACs.[7]

[7]Some treasurers are also informative in this regard, but they are not among the top 50 most important model features. This likely results from the low number of treasurers included in model estimation. Future iterations will expand the set of treasurers included.

Table 8: Top 50 Model Features by Variable Importance

Index	Mean Decrease Gini	Feature Type	Feature Name
1	4.478	aggregate donor attribute	ave. itemization ratio
2	3.528	individual vendor	American Technology Services
3	3.086	individual vendor	Unified Data Services
4	2.512	% expenditure category (FEC)	Solicitation And Fundraising Expenses
5	2.358	% expenditure category (CRP)	Fundraising
6	1.583	individual donor	Stanford, Charles
7	1.24	individual donor	Fink, Raymond N
8	1.221	aggregate donor attribute	% retirees
9	1.197	aggregate PAC attribute	ave. total fundraising ($)
10	1.169	individual vendor	Pitney Bowes
11	1.124	aggregate donor attribute	median contributor CFscore
12	0.945	individual donor	Berry, Yvonne
13	0.827	individual donor	Roberts, Dorothy B
14	0.815	% expenditure category (CRP)	Media
15	0.791	individual vendor	Ignite Payments
16	0.778	individual donor	Miller, Michael
17	0.726	aggregate donor attribute	ave. contributor CFscore
18	0.658	individual donor	Elliott, Donald G
19	0.633	% expenditure category (CRP)	Administrative
20	0.622	individual vendor	Blank Rome Llp
21	0.603	% expenditure category (FEC)	Administrative Salary Overhead Expenses
22	0.58	% expenditure category (CRP)	Unclassifiable
23	0.534	individual donor	Clark, Elloine
24	0.52	individual donor	Thomas, Alan B
25	0.508	individual donor	Mcdonald, Barbara
26	0.488	individual vendor	United States Postal Service
27	0.478	individual donor	Joiner, Mary F
28	0.444	% expenditure category (CRP)	Unclassified
29	0.427	individual vendor	Fedex
30	0.424	% expenditure category (CRP)	Campaign Expenses
31	0.416	individual donor	Mitchell, William E
32	0.408	individual vendor	Active Engagement
33	0.401	% expenditure category (FEC)	Unclassified
34	0.389	individual donor	Peabody, George
35	0.382	aggregate PAC attribute	active in 2014
36	0.376	individual donor	Edwards, Nick
37	0.369	individual donor	Johnson, Eric
38	0.365	individual donor	Nielson, Marilyn
39	0.36	individual donor	Mckibben, Lydia
40	0.355	% expenditure category (CRP)	Salaries
41	0.354	individual donor	Shaw, Robert
42	0.351	aggregate PAC attribute	active in 2012
43	0.344	individual donor	Binder, Adele
44	0.332	individual vendor	Huckaby Davis Lisker
45	0.32	individual donor	Birck, Katherine
46	0.32	individual donor	Tracy, P J
47	0.307	aggregate PAC attribute	active in 2016
48	0.293	individual donor	Locke, Gary
49	0.289	individual donor	Keller, David J
50	0.288	individual donor	Mcmanus, Jim

5 Conclusion

The proliferation of scam PACs in U.S. federal elections undermines the candidates and causes championed by campaign donors who fall victim to scam PACs, generates negative externalities in the political fundraising, and exacerbates inequality in campaign finance as a means of political participation. As is the case of most lemons problems, scam PACs thrive in information asymmetry. To reduce the informational barriers donors face in discerning scam PACs, thereby ameliorating the principal-agent problem between donors and the PACs to which they entrust with their campaign contributions, I propose a big-data approach to identify scam PACs. To this end, I start by quantitatively assessing a variety of observable attributes that appear to differentiate scam PACs from non-scam PACs, such as PACs' itemization ratio in fundraising, their budget allocation across expenditure categories, and their donor, treasurer, and vendor networks. Next, based on these descriptive findings, I construct a supervised algorithm that predicts PACs' likelihood of being scam PACs. Initial results from model estimation demonstrate the promise of supervised machine learning in helping donors distinguish scam PACs from non-scam PACs at scale.

As I improve upon my existing supervised algorithm, I will investigate several potential areas of improvement for data collection, model training, and model estimation. In terms of data collection, I will continue to expand the list of PACs that have been alleged as scam PACs by government or media reports. In addition, because of the controversial nature of referring to a PAC as a "scam PAC", I will document evidence of apparent financial self-dealing for each alleged scam PAC in my sample. Additionally, I will incorporate publicly available records of itemized contributions and PAC expenditures for all available election cycles (i.e., beyond the post-2010 cycles that my current empirical analysis is focused on).

In terms of model training, I will complete the process of standardizing identifiers for

PAC treasurers and vendors, which are key model features. Furthermore, I will refine my current measurement of target variables. Specifically, some of the non-scam PACs in my current data sample may in fact be scam PACs that have managed to evade public scrutiny, and such false negatives could bias my supervised algorithm. To alleviate this concern, I will consider identifying a subset of non-scam PACs as "legitimate PACs" (i.e., those whose fundraising and expenditure practices as well as their personnel networks leave little possibility for allegations of fraudulent conduct), and use only these PACs as the comparison set to scam PACs in model training.

In terms of model training, I will increase the memory limit currently imposed on the estimation process so as to include more model features that could be potentially informative, but are excluded at the moment due to data sparsity. Moreover, because scam PACs are currently rare relative to non-scam PACs in my data sample (although scam PACs raise much more money on average), I will adopt best practices for rare event detection in supervised machine learning beyond those I already employ in the paper. Last but not least, given the potentially sensitive label of "scam PACs" and the lack of a clear legal definition for them, I will investigate estimation approaches that penalize false positives more severely than false negatives.

Ultimately, I hope to use a more refined supervised algorithm for scam PAC detection to springboard future research on principal-agent problems in campaign fundraising. Specifically, I hypothesize that equipping campaign donors with informational tools to discern scam PACs, such as my supervised algorithm, could simultaneously reduce donors' contributions to scam PACs and maintain, if not increase, their contributions to candidate campaigns or legitimate PACs that scam PACs directly compete with (i.e., by claiming shared political objectives). In contrast, simply informing donors that they have fallen victims to scam PACs, by making them aware of the lemons problem without providing a means to overcome information asymmetry, could lead donors to withdraw from making campaign contributions altogether due to their inability to discern scam PACs

(Severns and Willis 2019). I envision a set of field experiments in which I administer these different information treatments to itemized donors who have contributed to scam PACs, and empirically verify the hypothesized changes in donor behavior in response to each type of information treatment. Results from such experiments could not only assess the efficacy of my supervised algorithm as a means to combat the problem of scam PACs, but also illuminate how information affects competition in the political marketplace and shapes the principal-agent relationships between campaign donors and recipients.

References

Ansolabehere, Stephen, John M. de Figueiredo, and James M. Snyder. 2003. "Why is There so Little Money in U.S. Politics?" *Journal of Economic Perspectives* 17(1): 105–130.

Arnsdorf, Isaac and Kenneth P. Vogel. 2016. "Trump Backers Face 'Scam PAC' Charges." *Politico*, May 16. https://www.politico.com/story/2016/05/scammers-feast-of-trump-fundraising-disarray-223141?cmpid=sf (Sep 1, 2020).

Bartels, Larry M. 2012. *Unequal Democracy: The Political Economy of the New Gilded Age*. Princeton: Princeton University Press.

Barber, Michael J., Brandice Canes-Wrone, and Sharece Thrower. 2017. "Ideologically Sophisticated Donors: Which Candidates Do Individual Contributors Finance?" *American Journal of Political Science* 61(2): 271–288.

Bonica, Adam. 2014. "Mapping the Ideological Marketplace." *American Journal of Political Science* 58(2): 367–386.

Bonica, Adam. 2018. "Inferring Roll-Call Scores from Campaign Contributions Using Supervised Machine Learning." *American Journal of Political Science* 62(4):830–848.

Bonica, Adam. 2019. Database on Ideology, Money in Politics, and Elections: Public version 3.0 [Computer file]. Stanford, CA: Stanford University Libraries. https://data.stanford.edu/dime

Bonica, Adam and Zhao Li. 2019. "Inferring Candidates' Issue-Specific Positions from Itemized Campaign Contributions Using Supervised Machine Learning." Working paper. https://www.dropbox.com/s/9b0g9fts8vtcvwl/Bonica%20and%20Li%20%282019%29.pdf?dl=0

Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45(1):5–32.

Drutman, Lee. 2015. *The Business of America is Lobbying: How Corporations Become Politicized and Politics Become More Corporate*. Oxford: Oxford University Press.

Federal Election Commission. 2017. *Registering as an SSF*. https://www.fec.gov/help-candidates-and-committees/registering-ssf/ (Sep 1, 2020).

Federal Election Commission. 2019a. *Understanding Nonconnected PACs*. https://www.fec.gov/help-candidates-and-committees/registering-pac/understanding-nonconnected-pacs/ (Sep 1, 2020).

Federal Election Commission. 2019b. *Statistical Summary of 24-Month Campaign Activity of the 2017-2018 Cycle*. https://www.fec.gov/updates/statistical-summary-24-month-campaign-activity-2017-2018-cycle/ (Sep 5, 2020).

Ferejohn, John. 1986. "Incumbent Performance and Electoral Control." *Public Choice* 50: 5–25.

Furnas, Alexander C. 2019. "Biasing Their Bosses: Staff Ideology, Motivated Reasoning, and the Distortion of Information in Congress." Working paper. https://static1.squarespace.com/static/59e25521b7411c07ef1410fa/t/5e1b8d543d0fa72e53e99a7b/1578863957461/BiasingBosses81819.pdf

Furnas, Alexander C., Timothy LaPira, Alexander Hertel-Fernandez, Lee Drutman, and Kevin Kosar. 2019. "Moneyed Interests, Information, and Action in Congress: A Survey Experiment." Working paper. https://static1.squarespace.com/static/59e25521b7411c07ef1410fa/t/5d5c6270b170bf00019fd581/1566335605396/Moneyed_Interests__Information__and_Action_in_Congress__A_Survey_Experiment.pdf

Graham, David A. 2019. "Political Fundraising Has a Big, Nasty Secret."*Atlantic*, July 29. https://www.theatlantic.com/ideas/archive/2019/07/conundrum-regulating-scam-pacs/594898/?utm_source=twitter&utm_medium=social&utm_campaign=share (Sep 1, 2020).

Grumbach, Jacob B. and Alexander Sahn. 2020. "Race and Representation in Campaign Finance." *American Political Science Review* 114(1): 206–221.

Grumbach, Jacob B., Alexander Sahn, and Sarah Staszak. "Gender, Race, and Intersectionality in Campaign Finance." Forthcoming at *Political Behavior*.

He, Haibo and Yunqian Ma. 2013. *Imbalanced Learning: Foundations, Algorithms, and Applications*, 1st ed. Hoboken, NJ: Wiley-IEEE Press.

Hertel-Fernandez, Alexander, Matto Mildenberger, and Leah Stokes. 2019. "Legislative Staff and Representation in Congress." *American Political Science Review* 113(1): 1–18.

Hirsch, Alexander V., Karam Kang, B. Pablo Montagnes, and Hye Young You. 2020. "Lobbyists as Gatekeepers: Theory and Evidence." Working paper. https://hyeyoungyou.files.wordpress.com/2020/06/lobbyists_as_gatekeepers.pdf

Hosmer, David W., Jr., Stanley Lemeshow, and Rodney X., Sturvidant. 2013. *Applied Logistic Regression*. Hoboken, NJ: John Wiley & Sons.

Hunter, Caroline C., Ellen L. Weintraub, Matthew S. Petersen, and Steven T. Walther.

2018. *Legislative Recommendations of the Federal Election Commission2018*. https://www.fec.gov/resources/cms-content/documents/legrec2018.pdf (Sep 1, 2020).

Janetsky, Megan. 2018. "Scam PACs Line Pockets by Misleading Donors."*OpenSecrets*, April 26. https://www.opensecrets.org/news/2018/04/scam-pacs-misleading-donors/ (Sep 1, 2020).

Kalla, Joshua L. and David E. Broockman. 2016. "Campaign Contributions Facilitate Access to Congressional Officials: A Randomized Field Experiment." *American Journal of Political Science* 60(3): 545–558.

Kleiner, Sarah. 2017. "Charities Employ Controversial Telemarketers to Tug on Heartstrings - and Loosen Purse Strings."*Center for Public Integrity*, Dec 13. https://publicintegrity.org/politics/veterans-charities/charities-employ-controversial-telemarketers-to-tug-on-heartstrings-and-loosen-purse-strings (Sep 1, 2020).

Kleiner, Sarah and Chris Zubak-Skees. 2019. "They Donated to Kids with Cancer. A Vegas Telemarketer Cashed in."*Tampa Bay Times*, Oct 30. https://www.tampabay.com/investigations/2019/09/12/they-donated-to-kids-with-cancer-a-vegas-telemarketer-cashed-in/ (Sep 1, 2020).

Kuhn, Max. 2008. "Building Predictive Models in R Using the Caret Package." *Journal of Statistical Software* 28(5): 1–26.

Lessig, Lawrence. 2011. *Republic, Lost: How Money Corrupts Congress–and a Plan to Stop It*. New York: Hachette Book Group.

Li, Zhao. 2018. "How Internal Constraints Shape Interest Group Activities: Evidence from Access-seeking PACs." *American Political Science Review* 112(4): 792–808.

Limbocker, Scott and Hye Young You. 2020. "Campaign Styles: Persistence in Campaign Resource Allocation." *Electoral Studies* 65:102140.

Lipton, Eric and Jennifer Steinhauser. 2015. "'Fire Paul Ryan'? Rebel PACs Hit Republicans, and It Pays."*New York Times*, Oct 23. https://www.nytimes.com/2015/10/24/us/politics/conservative-pacs-turn-attack-on-gop-leaders-into-fund-raising-tool.html?_r=0&auth=login-email&login=email&smid=tw-share (Sep 1, 2020).

Martin, Gregory J. and Zachary Peskowitz. 2015. "Parties and Electoral Performance in the Marketplace for Political Consultants." *Legislative Studies Quarterly* 40(3): 441–470.

Martin, Gregory J. and Zachary Peskowitz. 2018. "Agency Problems in Political Campaigns: Media Buying and Consulting." *American Political Science Review* 112(2): 231–248.

Min, Geeyoung and Hye Young You. 2019. "Active Firms and Active Shareholders: Corporate Political Activity and Shareholder Proposals." *Journal of Empirical Legal Studies* 48(1): 81–116.

Nyhan, Brendan and Jacob M. Montgomery. 2015. "Connecting the Candidates: Consultant Networks and the Diffusion of Campaign Strategy in American Congressional Elections." *American Journal of Political Science* 59(2): 292–308.

Persson, Torsten, Gerard Roland, and Guido Tabellini. 1997. "Separations of Power and Political Accountability." *Quarterly Journal of Economics* 112(4): 1163–1202.

Raymer, Matthew S. 2016. "Fraudulent Political Fundraising in the Age of Super PACs." *Syracuse Law Review* 66: 239–272.

Renshaw, Jarrett and Joseph Tanfani. 2020. "'Scam PAC' Fundraisers Reap Millions in the Name of Heart-Tugging Causes."*Reuters*, Jan 29. https://www.reuters.com/investigates/special-report/usa-fundraisers-scampacs/ (Sep 4, 2020).

Severns, Maggie and Derek Willis. 2019. "How Conservative Operatives Steered Millions in PAC Donations to Themselves."*Politico*, July 30. https://www.politico.com/story/2019/07/26/conservative-majority-fund-political-fundraising-pac-kelley-rogers-1428260 (Sep 1, 2020).

Weintraub, Ellen L. and Ann M. Ravel. 2016. *Proposal to Attack Scam PACs*. https://www.fec.gov/resources/about-fec/commissioners/weintraub/statements/2016-09_Memo--Scam-PACs.pdf (Sep 1, 2020).