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

Zhao Li

March 25, 2021

Preliminary Draft. Comments welcome.

Click here for the latest version.

'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 alleged scam PACs collectively raised more than $106 million, which could have fully funded 140 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 legitimate PACs on solicitation strategies, 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, Chuck Cameron, John Ferejohn, Andy Guess, Lewis Kornhauser, Tim LaPira, Greg Martin, Andrew Mayersohn, and participants at the Princeton Research in Experimental Social Science Workshop, the Annual Conference of the American Political Science Association, and the Colloquium on Law, Economics and Politics at New York University. All errors are my own.

1 Introduction

Principal-agent problems are ubiquitous in politics. 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 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, existing research has yet to explore an important form of principal-agent problems related to intermediaries in politics–the rise of so-called "scam PACs"–except for for a recent working paper on manipulation tactics employed in campaign emails (Mathur et al. 2020).[1] Scam PACs refer to non-connected PACs (i.e., political action committees that are not authorized by political candidates or parties, or sponsored by corporations or unions) that claim to support certain candidates or politi-

[1] A Google Scholar search for "scam PAC" or "scam PACs" returns only 14 results. Except for Mathur et al. (2020) and this paper, none of the remaining results are from political science, and all appear to focus on qualitative accounts of scam PACs.

cal causes when soliciting contributions, and yet redirect the money raised to enrich PAC treasurers, vendors, and other associates (Federal Election Commission 2019a). 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, based on my analysis, in the 2018 federal election cycle alone 84 PACs that have been alleged to be scam PACs by news outlets collectively raised $\$106,700,951$ in campaign contributions, which could have funded more than 140 average House campaigns in the same cycle.[2] In contrast to the lack of attention dedicated to this phenomenon, the Federal Election Commission has released memoranda highlighting the harms of scam PACs to donors, candidates, and legitimate PACs (Hunter et al. 2018; Weintraub and Ravel 2016). Similarly, the Federal Bureau of Investigation issued public warnings about scam PACs as a form of election crimes (Federal Bureau of Investigation 2020).

A substantial degree of information asymmetry makes it challenging for donors to discern scam PACs. First, 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 et al. 2018), and recent court cases further

[2]The former figure is based on my data collection of scam PACs as detailed in Section 3 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).

deprived the Commission of its ability to combat scam PACs that misleading imply affiliations with candidate campaigns during solicitation (e.g., *Pursuing America's Greatness v. FEC*). Second, while all PACs are required to disclose itemized disbursements to the Federal Election Commission, which may help to distinguish scam PACs from legitimate PACs, donors are often unaware of such publicly available records or do not understand how to take advantage of these resources (Hunter et al. 2018; Weintraub and Ravel 2016). Third, even if donors familiarize themselves with these information provided by the Federal Election Commission on PAC expenditures, they may fail to distinguish 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 expenditure decisions as laying the necessary groundwork for their PACs (even if doing so may conveniently create opportunities for financial self-dealing) (Lipton and Steinhauser 2015; Severns and Willis 2019).

To reduce the type of information asymmetry in political fundraising that enables scam PACs to proliferate, and to shed light on scam PACs as an under-explored type of "lemons" problems that undermines PACs' accountability to campaign donors (Akerlof 1973), I provide a first attempt at helping campaign donors discern scam PACs. I start with descriptive analyses that compare scam PACs to legitimate PACs on a variety of observable attributes. Section 3 details how I construct my data sample for both types of PACs. Section 4 presents preliminary results on the different solicitation strategies that scam PACs and legitimate PACs appear to employ. Section 5 distinguishes scam PACs and legitimate PACs on various aspects of PAC fundraising and expenditure patterns, including aggregate patterns (e.g., fundraising and disbursement size, itemization ratio in fundraising, budget allocation across expense categories), donor characteristics (e.g., ideology and age), and PACs' networks of donors, treasurers, and vendors.

Building upon these descriptive findings, I then construct a supervised algorithm that predicts the likelihood of individual PACs being scam PACs based on publicly available campaign finance data. As shown in Section 6, my supervised algorithm attains a high level of out-of-sample predictive performance, and may be suitable as a tool for systematic detection of scam PACs in U.S. federal elections. Moreover, rather than relying on arbitrary and often imprecise rules of thumbs in classifying scam PACs (Janetsky 2018; Kleiner and Zubak-Skees 2019), this algorithm provides validated measures of which specific observable attributes of PACs are the more predictive of scam PACs at the margin, which could provide useful heuristics for donors seeking to discern scam PACs. Section 7 summarizes potentially fruitful areas of improvement for this project, and outlines field experiments that may build upon this paper in testing the effectiveness of information interventions in ameliorating the problem of scam PACs in the fundraising marketplace.

2 Literature Review and Motivation

2.1 The Conservative Majority Fund: A Case Study

While there currently exists no legal definition of what constitutes a "scam PAC," memoranda published by the Federal Election Commission conceptualize them as non-connected PACs that satisfy the following two criteria (Hunter et al. 2018; Weintraub and Ravel 2016). First, despite promises to prospective donors to support political candidates or causes through campaign contributions or independent expenditures, scam PACs tend to direct large portions of their expenditures to overhead costs (e.g., salaries, fundraising) that are unlikely to affect election outcomes or political discourse. Second, scam PACs' disbursements are often used as a conduit of self-enrichment for campaign consultants and vendors.

The Conservative Majority Fund, a scam PAC that was terminated in 2019, is a case in point. A chief operative behind the Conservative Majority Fund and a number of other scam PACs, Kelley Rogers is a seasoned political consultant who previously worked for the American Conservative Union, which organizes the annual Conservative Political Action Conference (CPAC). Rogers leveraged these work experiences to cultivate donor lists and perfect solicitation tactics that galvanized small-dollar, elderly donors, such as by purporting to lobby states to remove Barack Obama from the ballot and to pay for undercover exposes of the Obama campaign's attempt to commit voter fraud (Severns and Willis 2019). Under the management of Kelley Rogers as well as Scott B. MacKenzie, another campaign veteran who served as the official treasurer of the Conservative Majority Fund, the Fund raised almost \$10 million since 2012 and yet donated just $\$48,400$ to political candidates (Severns and Willis 2019). Among the remaining disbursements, MacKenzie received $\$172,000$ in salaries as the PAC treasurer; Strategic Campaign Group, for which Rogers formerly served as the president, received $\$229,000$ in consulting fees; and millions of dollars were reported as campaign or media expenses to the Federal Election Commission when they were in fact used for fundraising calls and media consulting work that produced little tangible output of mass communication (Severns and Willis 2019). For their conduct related to the Conservative Majority Fund and other scam PACs, Rogers and MacKenzie were eventually sentenced to federal prison for committing wire fraud by "[swindling] millions of dollars from individuals attempting to participate in our democratic process," and for making false statements to the Federal Election Commission and creating fake invoices in order to mask their self-dealing activities (Stueve 2019, 2020).

2.2 Scam PACs as a "Lemons" Problem in the Political Marketplace

Scam PACs can be conceived as a form 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 candidates' accountability to voters (i.e., can campaign contributions corrupt elect officials and erode the representation of constituent interests?) (e.g., Bartels 2012; Lessig 2011), candidates' 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), or political intermediaries' accountability to candidates (do campaign vendors efficiently serve their client candidates?) (e.g., Martin and Peskowitz 2015, 2018). 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 (Center for Responsive Politics 2021).

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 intend to 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 may threaten to undo recent progress in the diversification of the donor pool and increase inequality in participation in campaign finance (e.g., Grumbach and Sahn 2020; Grumbach, Sahn, and Staszak n.d.).

2.3 *Information as a Governance Mechanism*

While Kelley Rogers and Scott B. MazKenzie faced prison time for their conduct in connection to the Conservative Majority Fund, it was a rare instance of prosecution. According to the Federal Election Commission, the Commission has no authority to regulate fraudulent campaign fundraising conduct (other than punishing false disclosure to the Commission) unless Congress amends existing campaign finance regulations (Hunter et al. 2018; Weintraub and Ravel 2016). Absent such legislative reforms, the Commission argues that the next best alternative would be to provide prospective donors with more information about the fundraising and disbursements of PACs, such as using publicly available records on the Commission's website to construct ratings of PACs similar to those issued for non-profit organizations by charitable watchdogs (Hunter et al. 2018; Weintraub and Ravel 2016). Charity Navigator, for example, assesses non-profit groups on financial health metrics as well as accountability and transparency metrics based on Form 990's filed by these groups (Charity Navigator 2021). Moreover, existing research finds that organizations that received higher ratings from charitable watchdogs subsequently received more donations, suggesting that these ratings may fill an information gap that prospective charitable donors face and enhance accountability and trust within non-profit spheres (Gordon, Knock, and Neely 2009; Yoruk 2016).

However, replicating the works of charitable watchdogs directly using publicly available data provided by the Federal Election Commission is challenging for at least two reasons. First, while PACs have to disclose itemized disbursements, PACs, unlike non-profit organizations, are not currently required to provide any information with regard to the potential of conflicts of interest or other aspects of PAC governance structures. Second, watchdogs such as Charity Navigator rate non-profit organizations based on a number of specific formulas and cutoffs that require varying degrees of discretion. For instance, a general non-profit group receives a financial efficiency of 10 if it spends 85% or more of its budget on programs and services they exist to provide, 0 if less than 50%, and a intermediate score of $\frac{10\times(RawScore-0.5)}{0.35}$ if this expenditure share (i.e., the raw score) is between 50% and 85% (Charity Navigator 2021). In the absence of guidelines from the Federal Election Commission or validation exercises, it is hard to construct arbitrary formulas using campaign finance disclosure data that accurately identify scam PACs. For example, in one of its investigative reports on scam PACs, the Center for Public Integrity focused on PACs that raised at least $\$10,000$ during at least one election cycle, and classified them as scam PACs if they received more than 50% of its total fundraising from unitemized donors (i.e., those giving less than $200 in a calendar year) and spent more than 50% of their total expenditures on fundraising, wages, and administration (Kleiner and Zubak-Skees 2019). However, when compared against the set of scam PACs that I have collected based on news reports (see Section 3 for detail), the Center for Public Integrity's methodology for identifying scam PACs produces a high false positive rate of 64% (i.e., 56 out of 87 predicted scam PACs); Table 1 displays the full confusion matrix. This example illustrates the potential danger of relying on intuitive rules of thumb to classifying PACs as potential scam PACs given the sensitivity of such a label.

Table 1: Confusion Matrix of Scam PAC Predictions Based on the Center for Public Integrity's Methodology

Prediction \ Reference	False	True
False	1428	49
True	56	31

In lieu of arbitrary and often subjective criteria for identifying scam PACs, I propose to first conduct a quantitative assessment of observable attributes that appear to distinguish PACs that have been alleged as scam PACs by news reports from PACs without such allegations, and then use publicly available records on PACs to construct 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 legitimate PACs based on publicly available records of PACs as urged by the Federal Election Commission (Weintraub and Ravel 2016). Moreover, supervised algorithms have a number of important advantages. First, supervised models infer from data, rather than assuming, the unobserved and potentially complex mappings between PACs' characteristics and their probabilities of being scam PACs. Second, supervised machine learning also permits us to incorporate a large amount of potentially useful information about PACs (e.g., PACs' donor and personnel networks) into the model estimation and prediction processes. Third, insofar as the estimated algorithms demonstrate high out-of-sample predictive performance, they 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 news reports. Fourth, feature important results from supervised algorithms can provide validated measures of observable traits that are indicative of scam PACs. In short, supervised algorithms for scam PAC detection may ameliorate information asymmetry in the political marketplace, thereby restoring accountability and trust in campaign fundraising (Hunter et al. 2018; Weintraub and Ravel 2016).

3 Data Construction

3.1 Data sources and time frame

I collect publicly available federal campaign finance records from the following three sources: the Federal Election Commission (FEC)'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 primarily active in the past decade, and that the 2020 data are not yet fully available online.

In addition, for exploratory analysis of PACs' solicitation strategies, I collect for a select number of PACs their Facebook advertisements through the Facebook Ad Library's API, and their public Facebook posts through CrowdTangle. The Facebook Ad Library makes all political advertisements (broadly defined by Facebook and includes issue as well as campaign advertising) posted since May 2018 available (Fowler et al. 2020). CrowdTangle provides data on all public posts that remain on public Facebook pages that are verified or have more than $100,000$ followers (CrowdTangle Team 2020). For both of these sources, I collect all available data starting from the earliest dates provided to present day.

3.2 Identifying scam PACs

To collect a sample of scam PACs, I scrapped the search results for keywords "scam PAC" or "scam PACs" on Google News that were published between January 1, 2011 and December 31, 2020. Within these search results, I retained those that named specific PACs that were alleged to be scam PACs either by the authors of the news articles or the sources they cited. The remaining 288 news articles mentioned 99 unique alleged scam PACs. In my analyses, I focus on 84 of these alleged scam PACs that were active some time during the $2010 - 2018$ election cycles, and raised at least \$$10,000$ per active cycle. For the remainder of this paper, I refer to these alleged scam PACs as "scam PACs" for simplicity, though it is worth stressing that there is currently no legal definition of such PACs.

Table 2 lists each of the scam PACs identified using the aforementioned procedure, the number of unique news articles that mentioned each scam PAC, the news outlets that mentioned each scam PAC, and the inferred ideological leaning of each scam PAC based on whether its recipient CFscore is above or below zero (Bonica 2014). Many conservative-leaning scam PACs seek to attach themselves to the Tea Party movement (e.g., the Tea Party Leadership Fund), Donald Trump (e..g., Patriots for Trump), or other prominent conservative political figures such as Ben Carson (e.g., National Draft Ben Carson PAC). In addition, a much smaller number of liberal-leaning scam PACs claim to support prominent progressive Democrats such as Alexandria Ocasio-Cortez (e.g., Justice Democrats) and Bernie Sanders (e.g., Americans Socially United).

The fact that the overwhelming majority (82.1%) of scam PACs in Table 2 appears to be conservative-leaning may raise concerns about potential bias in news coverage of scam PACs. However, several observations ameliorate such concerns. First, my data collection method, which relies on keyword searches on Google News, should not preclude media sources on any side of the political spectrum, including right-leaning media outlets such as the Daily Wire, National Review, and Fox News. Second, it is ex ante unclear how partisan bias would skew coverage of scam PACs. On the one hand, outlets that prefer one party may be more inclined to report on scam PACs of the opposite partisan persuasion in an attempt to shame partisan opponents, and such incentives might partially account for the dominance of conservative scam PACs in my sample insofar as media outlets tend to lean left (Groseclose and Milyo 2005). On the other hand, there may be incentives to prioritize coverage of scam PACs that share an outlet's partisan preference in order to alert potential donors and limit the damage these scam PACs may cause to an outlet's preferred party. For example, Erick Erickson, the editor of RedState.com, called scam PACs "a blight on the GOP" (Altman and Scherer 2014). Similarly, the National Review bemoaned scam PACs as "the Right's Grifter Problem" (Geraghty 2019). Third, many Republican politicians have condemned the prevalence of conservative scam PACs that target their supporters, including and not limited to Donald Trump (Arnsdorf and Vogel 2016), Mike Pence (Lewis 2015), Trey Gowdy (Lipton and Steinhauser 2015), Ken Cuccinelli (Lewis 2015), and Alan West (Janetsky 2018).

Table 2: A List of Alleged Scam PACs

PAC Name	Number of unique news articles	Source(s)	Inferred Ideology

Conservative Strikeforce	20	Politico, Center for Responsive Politics, The Hill, Washington Post, Campaigns and Elections, Consultancy.uk, The Salt Lake Tribune, National Review, ProPublica, Center for Public Integrity, Daily Wire	conservative
Conservative Majority Fund	19	Washington Post, Campaigns and Elections, Center for Responsive Politics, Politico, The Atlantic, National Review, ProPublica, Center for Public Integrity, Daily Wire	conservative
Republican Majority Campaign	10	Politico, Center for Responsive Politics, Campaigns and Elections, WUWM 89.7 FM Milwaukee Public Radio, Lifehacker, National Review, Salon, floridabulldog.org	conservative
Tea Party Majority Fund	10	Politico, Center for Responsive Politics, National Review, Campaigns and Elections, Washington Post, ProPublica, Center for Public Integrity, Daily Wire	conservative
U.S. Virgin Island Republican Party (VIGOP)	8	Center for Responsive Politics, Politico, St. Croix Source, The Salt Lake Tribune, Center for Public Integrity	conservative
Americans for Law Enforcement	7	Campaigns and Elections, ABC 15 Arizona, WUWM 89.7 FM Milwaukee Public Radio, National Review, Salon, floridabulldog.org	conservative
Americans for the Cure of Breast Cancer	7	Politico, The Cap Times, Center for Public Integrity, ProPublica, Reuters, The Week	conservative

Bold Conservatives (F.K.A. Draft Sherriff David Clarke for U.S. Senate)	6	Milwakee-Wisconsin Journal Sentinel, Milwaukee Journal Sentinel, Roll Call, WGBH (Boston Public Radio), Center for Responsive Politics	conservative
Great America PAC	5	Politico, The Daily Beast, National Review, The Week	conservative
The Police Officers Support Association	5	CNN, Salon	conservative
Life and Liberty PAC	5	ABC 15 Arizona, WUWM 89.7 FM Milwaukee Public Radio, Salon, floridabulldog.org	conservative
Tea Party Leadership Fund	4	Politico, New York Times, Snopes	conservative
Restore American Freedom and Liberty	4	Politico, Milwakee-Wisconsin Journal Sentinel, Metro Weekly, Blue Virginia	conservative
Freedom's Defense Fund	4	Center for Responsive Politics, Campaigns and Elections	conservative
Conservative Action Fund	4	Politico, Washingtonian, New York Times	conservative
Patriots for Economic Freedom	3	Politico, Blue Virginia	conservative
Voter Education PAC	3	The Arizona Republic, ABC 15 Arizona, WUWM 89.7 FM Milwaukee Public Radio	conservative
Grassroots Awareness PAC	3	Campaigns and Elections, ABC 15 Arizona, WUWM 89.7 FM Milwaukee Public Radio	conservative
Standing by Veterans	3	Politico, The Cap Times, ProPublica	conservative

Association for Emergency Responders and Firefighters	3	Politico, The Cap Times, ProPublica	conservative
Firefighters and Emergency Responders Coalition	3	Reuters, The Week	conservative
Protect Our Future	3	ABC 15 Arizona, WUWM 89.7 FM Milwaukee Public Radio, Salon	conservative
Heroes United	3	AARP, floridabulldog.org, Salon	conservative
US Veterans Assistance Foundation	3	Politico, The Cap Times, ProPublica	conservative
Cops and Kids Together	3	Politico, ProPublica	conservative
Support American Leaders	3	Politico, KCBS (CBS Radio News), The Daily Beast	unknown
National Campaign PAC	3	The Arizona Republic, ABC 15 Arizona, WUWM 89.7 FM Milwaukee Public Radio	conservative
RightMarch.com	3	ABC 15 Arizona, WUWM 89.7 FM Milwaukee Public Radio, National Review	conservative
Americans Socially United	2	Center for Public Integrity, The Week	liberal
Democratic Coalition Against Trump	2	The Daily Beast, Splinter News	liberal

Committee to Restore America's Greatness	2	Politico, National Review	conservative
National Draft Ben Carson PAC	2	WGBH (Boston Public Radio), The Dispatch	conservative
Brand New Congress	2	Capital Research Center, NBC News	liberal
American Horizons	2	Politico	conservative
Action Coalition	2	ABC 15 Arizona, WUWM 89.7 FM Milwaukee Public Radio	conservative
Americans for Police and Trooper Safety	2	Politico, ProPublica	conservative
Black Americans to Re-Elect the President	2	Fox News, The Dispatch	conservative
Justice Democrats	2	Capital Research Center, NBC News	liberal
BAMPAC	2	Politico, Roll Call	conservative
Stop Hillary PAC	2	Politico, Milwakee-Wisconsin Journal Sentinel	conservative
Put Vets First	2	Center for Responsive Politics, National Review	conservative
Campaign to Defeat Barack Obama	2	Politico, The Grand Forks Herald	conservative
Justice-PAC	1	Politico	conservative
Patriot Super PAC	1	Politico	conservative
Make America Great Again	1	Politico	conservative
BlakPAC	1	Milwakee-Wisconsin Journal Sentinel	conservative

Go Big Go Bold	1	Center for Responsive Politics	conservative
RallyPAC	1	CNBC	conservative
Patriots for Trump	1	Campaigns and Elections	conservative
Feel Bern	1	The Week	liberal
Conservative Freedom Fighters	1	Center for Responsive Politics	conservative
Amish PAC	1	WGBH (Boston Public Radio)	conservative
Tea Party Forward	1	Politico	conservative
Tea Party Patriots	1	New York Times	conservative
Americans for Progressive Action USA	1	Politico	conservative
Stop Pelosi PAC	1	Politico	conservative
Draft Newt PAC	1	Politico	conservative
Creative Majority (CMPAC)	1	Capital Research Center	liberal
People's House Project	1	McClatchy	liberal
National Send Them Packing Committee	1	WGBH (Boston Public Radio)	conservative
Breast Cancer Health Council	1	Fits News	conservative
MAGA Coalition	1	Politico	conservative
National Assistance Committee	1	Politico	unknown
United American Veterans	1	ProPublica	conservative

Bikers for the President	1	Politico	conservative
American Coalition for Injured Veterans	1	Reuters	conservative
Children's Leukemia Support Network	1	ProPublica	unknown
Firefighters Alliance of America	1	ProPublica	unknown
Heart Disease Network of America	1	ProPublica	unknown
Police Officers Defense Alliance	1	ProPublica	unknown
Autism Hear Us Now	1	ProPublica	liberal
America Fighting Back	1	Politico	conservative
The Fight	1	Center for Responsive Politics	liberal
United Veterans Alliance of America	1	ProPublica	conservative
Keeping America Great	1	Politico	conservative
Republican Majority Campaign	1	ABC 15 Arizona	conservative

The Madison Project	1	The Cap Times	conservative
Constitutional Rights PAC	1	New York Times	conservative
Black Republican PAC	1	Politico	conservative
Our Country Deserves Better	1	Politico	conservative
Western Representation PAC	1	Politico	conservative
One Nation	1	The Desert Sun	conservative
Combat Veterans for Congress	1	The Desert Sun	conservative
Coalition of Americans for Political Equality (CAPE)	1	Politico	conservative

For each scam PAC, I link it to its corresponding FEC committee ID's by PAC name, and compile data on its fundraising and disbursements, itemized campaign donors, and PAC treasurers and vendors from the FEC, CRP, and DIME. I identify all unique campaign donors (using *bonica.cid* in DIME as time-invariant donor ID's) who have made at least one itemized contribution to one or more scam PACs within the time frame of my analysis. In addition, I identify unique treasurers and vendors who have worked for scam PACs using their standardized names and reported zip codes, since existing expenditure data sources do not provide official identifiers for PAC personnel.

3.3 *Identifying comparable legitimate PACs*

Next, in order to construct valid comparison groups for scam PACs, I create a list of what I hereafter refer to as "legitimate PACs", i.e., PACs that have similar organizational structures as scam PACs and have not been alleged as scam PACs by any news articles that I scrapped. Specifically, since all scam PACs fall into the category of "non-connected PACs" as defined by the Federal Election Commission, each legitimate PAC in my sample must satisfy the same set of criteria, i.e., 1) it must not be authorized by any candidate campaigns; 2) it must not be an official party committee; 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 legitimate PACs in my sample have comparable fundraising and disbursement needs as scam PACs, including the lack of a restricted class of donors (e.g., firm employees or union members) and a natural tendency to have higher fractions of operating expenses (e.g., fundraising and administrative work) for organizational maintenance. In addition, I restrict attention to legitimate PACs that on average raised more than $10,000 during their active cycles. In total, between the 2010 and 2018 election cycles, there were 1,871 legitimate PACs per my definition, among which 774 are liberal leaning and 911 are conservative leaning based on whether their corresponding recipient CFscores are below or above zero (Bonica 2014). I also identify unique donors, treasurers, and vendors for legitimate PACs using the same procedures as described for scam PACs.

4 Exploratory Analyses of Solicitation Strategies

The case of the Conservative Majority Fund discussed earlier, along with media reports, suggest that scam PACs may differ from legitimate PACs in their solicitation strategies. For instance, scam PACs may be more likely to reference controversial political figures or issues and employ emotionally-charged language when communicating with prospective donors (Geraghty 2019; Severns and Willis 2019). Moreover, particularly for conservative scam PACs, elderly populations appear to be particularly receptive to scam PACs' rhetoric (Graham 2019; Severns and Willis 2019). While verifying these claims at scale can be challenging, not least because scam PACs may be more likely to leverage means of offline solicitation such as telemarketing and mailers (Severns and Willis 2019), comparisons of the solicitation strategies pursued by both scam PACs and legitimate PACs on shared platforms may nonetheless illuminate why scam PACs are able to siphon donations at the expense of candidate campaigns and legitimate PACs.

4.1 Use of Facebook advertisements versus Facebook posts

One such shared platform is Facebook. For the 84 scam PACs shown in Table 2, I manually link each of them to their affiliated Facebook page ID's where available, thereby allowing me to search for their advertisements via the Facebook Ad Library API, and public posts through CrowdTangle if their pages meet CrowdTangle's thresholds on verification status or follower counts. Given the large number of legitimate PACs in my sample, I have yet to extract all of their corresponding advertising or posting data. Instead, for now I manually linked a subset of the largest legitimate PACs to their Facebook records such that I have a roughly balanced number of scam PACs and legitimate PACs for whom such records are available. Given the incomplete nature of the sample, all results shown in this section are tentative and may change as more legitimate PACs are included in the analyses.

Tables 3 and 4 display the summary statistics associated with the use of each of these online means of communications across six categories of PACs: all scam PACs, all legitimate PACs, liberal-leaning scam PACs, liberal-leaning legitimate PACs, conservative-leaning scam PACs, and conservative-leaning legitimate PACs. Note that in Tables 3 the average number of Facebook advertisements run by each type of PACs as well as the average expenditures on Facebook advertisements by PAC type are calculated using only PACs that have advertising histories on Facebook. Similarly, in Table 3, the average years on Facebook for each type of PACs as well as the average number of Facebook posts for each type of PACs are calculated using only PACs with Facebook pages recorded by CrowdTangle's database.

Table 3: Summary Statistics for Facebook Ads

Statistic	Scam PACs (all)	Legitimate PACs (all)	Scam PACs (liberal)	Legitimate PACs (liberal)	Scam PACs (conservative)	Legitimate PACs (conservative)
No. Actively Advertising PACs	9	28	4	7	5	16
Ave. No. Facebook Ads	1,191	2,598	2,033	2,368	518	2,521
Ave. Expenditures on Facebook Ads	$ 176,612	$2,596,415	$ 197,624	$4,645,758	$ 159,801	$1,996,301

Table 4: Summary Statistics for Facebook Posts

Statistic	Scam PACs (all)	Legitimate PACs (all)	Scam PACs (liberal)	Legitimate PACs (liberal)	Scam PACs (conservative)	Legitimate PACs (conservative)
No. PACs with Large Facebook Pages	22	29	6	6	16	20
Ave. Years on Facebook	6.18	6.52	4.00	6.50	7.00	6.35
Ave. No. Facebook Posts	4,813	2,550	11,514	1,422	2,300	2,454

In Tables 3, it appears that even among scam PACs and legitimate PACs that advertised on Facebook, scam PACs ran fewer advertisements and spent far less money on them than legitimate PACs regardless of ideological leanings. Given the relatively low cost of Facebook advertisements compared to more traditional means of campaign or issue advertising via television or ratio (Fowler et al. 2020), this finding appears to be consistent with the narrative that scam PACs are less likely to engaging in political spending via mass media (Hunter et al. 2018; Weintraub and Ravel 2016).

However, among scam PACs and legitimate PACs with large Facebook pages, there is little noticeable difference in the frequency of usage of Facebook posts as shown in Table 3. In fact, liberal scam PACs in particular appear to post a lot more frequently than liberal legitimate PACs even though the former has generally been active on Facebook for a shorter amount of time. The contrast between Tables 3 and Tables 4 suggests that scam PACs may prefer to reach potential donors via free versus paid channels through Facebook. This does not, however, mean that scam PACs would never pay for engagement with potential donors. In fact, as will be shown in Section 5, a hallmark of conservative-leaning scam PACs is the high fraction of expenditures on fundraising. Facebook advertisements may be simply inefficient at reaching the target groups of prospective donors for scam PACs, and provide no opportunities for self-dealing among political consultants.

4.2 Audience Demographics for Facebook Advertisements

One of the useful features of the Facebook Ad Library API's database is that it provides data on the estimated age and gender patterns of Facebook users who saw a particular advertisement (Fowler et al. 2020). Even though such data are not the same as the demographic targeting strategies pursued by advertisers (Fowler et al. 2020), they may still allow a glimpse into how scam PACs' and legitimate PACs' audiences on Facebook may differ. For a scam PAC or a legitimate PAC with a history of Facebook advertisements, I calculate the average gender ratios over all of its previous advertisements, and I analogously compute the average share of audience belonging to each of the age brackets provided by Facebook. Figure 1 displays the result on gender, and Figure 2 on age.

Interestingly, while no news articles or government reports that I have come across on this topic suggested gender balances across prospective donors of scam PACs versus legitimate PACs, Figure 1 shows that the audience of liberal-leaning scam PACs' Facebook advertisements may skew more male than those for liberal-leaning legitimate PACs. In comparison, Facebook advertisements by conservative-leaning PACs display little gender difference in their audiences.

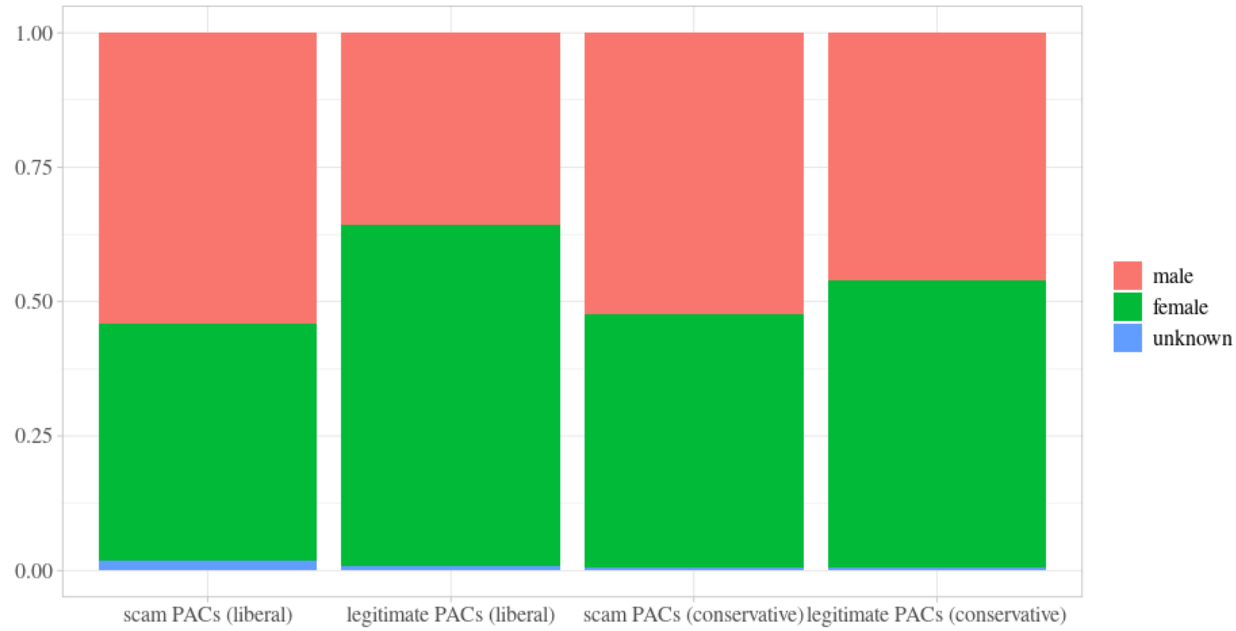

Figure 1: Gender

In terms of age distributions, Figure 2 appears to confirm a widespread claim that (conservative-leaning) scam PACs appear to attract older populations than legitimate PACs (Graham 2019; Lewis 2015; Severns and Willis 2019). Curiously, my later analyses using itemized contributors' self-disclosed retiree status replicates this apparent correlation between age and interest in conservative scam PACs (see Section 5), and also dovetails with research on the the age effect in the consumption and sharing of fake news and the relative lack of digital literacy (Guess, Nagler, and Tucker 2019; Guess, and Munger 2020). The link between my analyses on scam PACs and the political misinformation/disinformation literature is not a mere fluke. As the case study of the Conservative Majority Fund illustrates, and the remainder of this section shows, scam PACs may be more likely to employ the type of incendiary and sensationalized rhetoric–often seen in fake news (Tucker et al. 2018)–when attempting to solicit contributions.

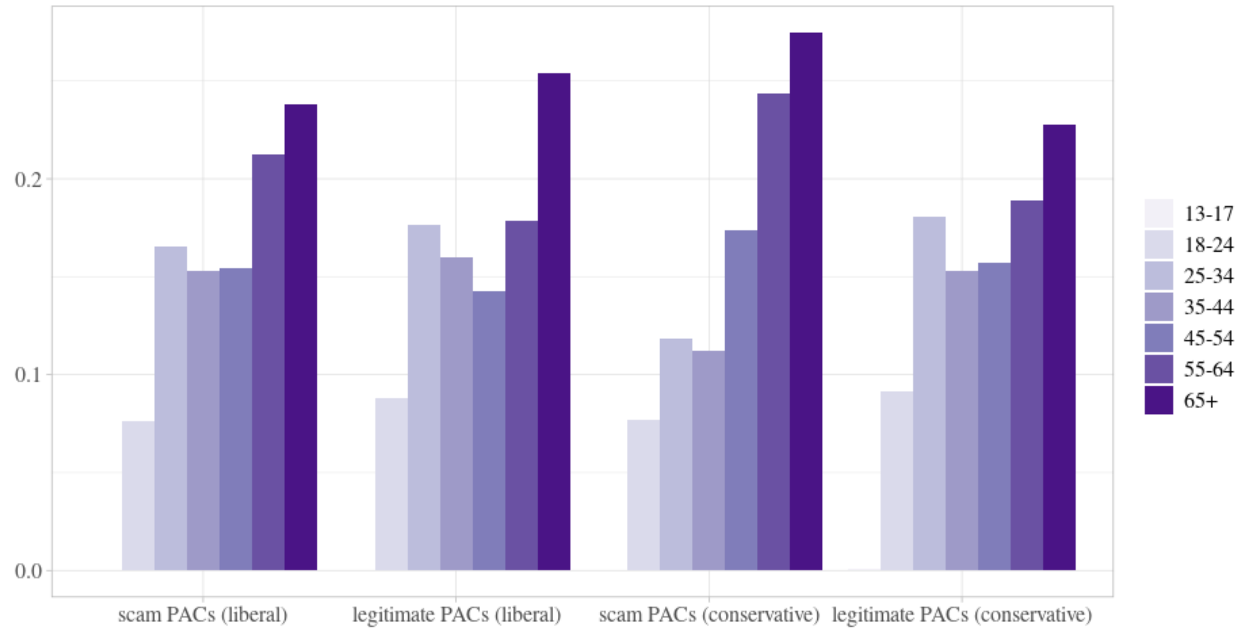

Figure 2: Age

4.3 *Sentiment Analysis*

I conducted preliminary sentiment analyses separately for Facebook advertisements and posts. To do so, I combine all texts shown within a given advertisement or a post, including the main bodies of text as well as any text transcribed from audiovisual data or linked sources. I then calculate, for each unique advertisement or post, the net share of all words in its text that are classified as exhibiting positive rather negative sentiment using Bing's sentiment lexicon. Departing from the default lexicon, I remove the word "Trump" as a positive word given its use as a name rather than a verb in most of my text data, although further pre-processing may still be needed.

I then aggregate these measures of message sentiment at the PAC level by averaging the net shares of positive words across all advertisements or all posts published by a given PAC. For Facebook advertisements, I currently find no statistically significant correlation between propensity to employ positive language and the likelihood of a PAC being classified as a scam PAC. This is true regardless of PACs' ideological orientation. This null result may stem from a relatively low sample of scam PACs that have advertised on Facebook (see Table 3), or it may reflect a genuine lack of distinction in the sentiment of advertisements. In contrast, for Facebook posts, the average net share of positive words by a conservative-leaning PAC is negatively associated with it being a conservative-leaning scam PAC; the corresponding Pearson correlation coefficient is -0.325 and statistically significant. Put differently, conservative scam PACs may be more inclined to use words associated with negative sentiment in their Facebook posts than legitimate PACs of the same ideological leaning. I do not find a precisely estimated correlation in either direction among liberal-leaning PACs for Facebook posts.

4.4 Topic Analysis

I also explored the topics mentioned in Facebook advertisements and posts by PACs. To this end, I once again combine all texts shown within a given advertisement or a post. I then estimate a separate topic model for each means of communication–advertisements versus posts–and each year use the *stm* package in R (Roberts, Stewart, and Tingley 2019). 60-topic models appear to reasonably summarize both types of text data, though further validation on the optimal number of topics is needed. Figure 3 displays the top 20 topics by prevalence for Facebook advertisements run by scam PACs or legitimate PACs in 2018, and Figure 4 displays analogous results for Facebook posts in 2018.

Within each topic model, I calculate the average frequency of each topic across all texts (i.e., advertisements or posts) by PAC, and test whether scam PACs are more or less noticeably likely to reference certain topics compared to legitimate PACs. While results here are preliminary, I only find a small number of topics that distinguish scam PACs from legitimate PACs for Facebook advertisements (see Figure 3). For example, within conservative-leaning PACs, scam PACs were much more likely to mention topic 37, which appears to be about threats of impeachment that Donald Trump faced in 2018. Additionally, compared to conservative-leaning legitimate PACs, conservative-leaning scam PACs discussed topic 4 more frequently, which seems to be a call for donations to aid Ted Cruz against an electoral challenge from Beto O'Rourke in 2018.

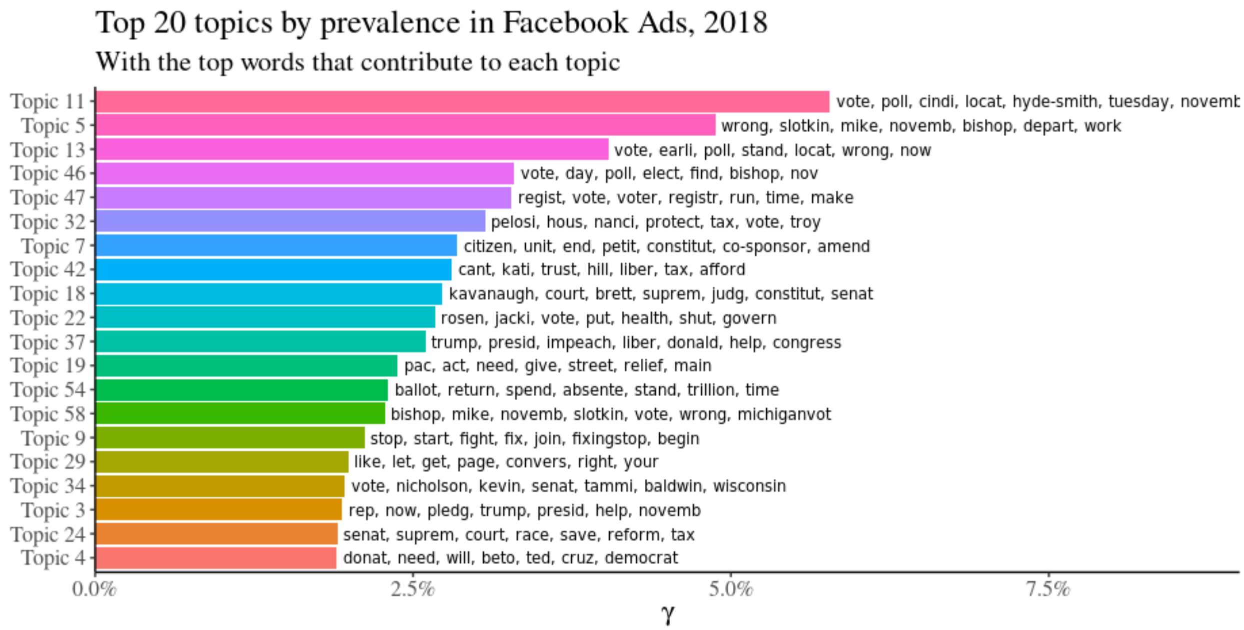

Figure 3: Topic Ads

In contrast to results from Facebook advertisements, scam PACs and legitimate PACs exhibit significantly different tendencies to engage in a range of topics in their Facebook posts (see Figure 4). Here are a non-exhaustive list of examples. Relative to legitimate PACs of comparable ideological persuasion, both liberal- and conservative- leaning scam PACs appeared to be more eager to bring up contentious social issues or political figures such as topic 54 (debates on gun control versus gun rights), topic 30 (threats of impeachment facing Donald trump), topic 25 (Alexandria Ocasio-Cortez and other Progressive Democrats), topic 18 (the Mueller investigation), topic 47 (the FBI's investigation into potential Russian interference in U.S. elections), and topic 35 (race relations and the use of force). Moreover, across the ideological spectrum, scam PACs were less likely to mention topics of what appears to be more bread-and-butter economic issues such as topics 55 and 27. Additionally, within conservative-leaning PACs, scam PACs posted much more often on topic 7 (sexual assault allegations against Brett Kavanaugh) on Facebook than did legitimate PACs, though liberal-leaning scam PACs and legitimate PACs were equally likely to engage in this topic.

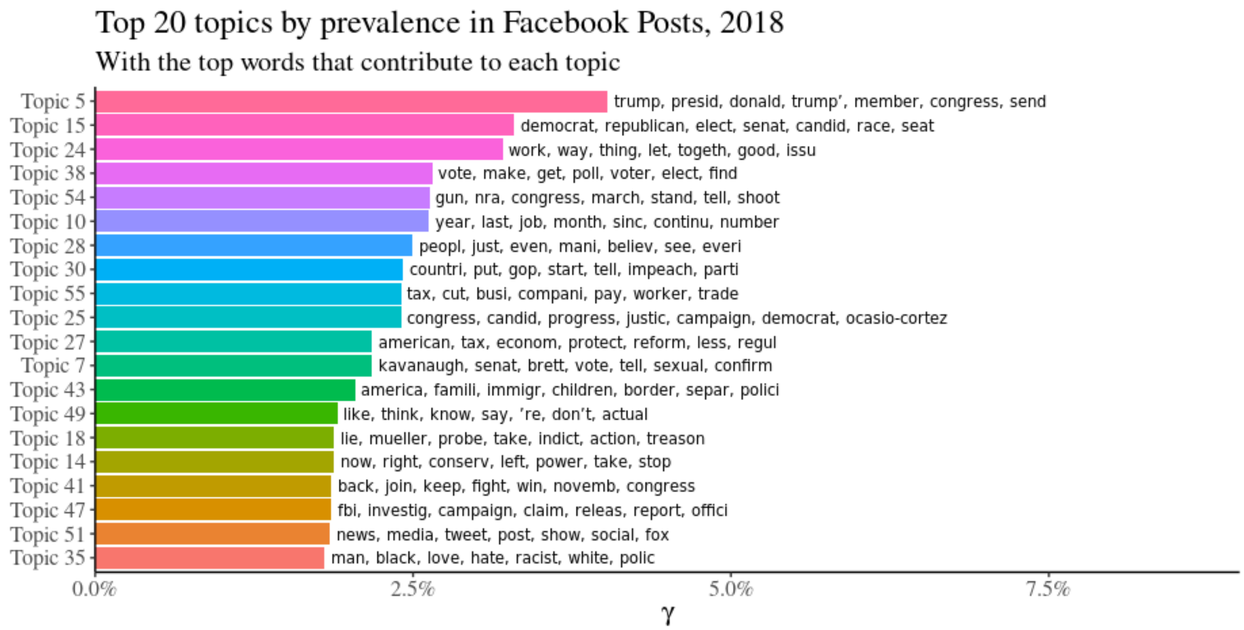

Figure 4: Topic Posts

5 Analyses of Campaign Finance Records

While the previous section provides tentative findings on differences in solicitation strategies pursued by scam PACs and legitimate PACs, this section highlights observable attributes that may be useful for discernment of scam PACs through supervised learning. To that end, I conduct a series of descriptive comparisons on PACs' aggregate fundraising and expenditure patterns, donor attributes, and personnel networks.

5.1 PAC fundraising and expenditures

To examine fundraising and expenditure patterns, I first compute average amounts of money raised and spent by PACs 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 legitimate PACs (Weintraub and Ravel 2016), I compute average itemization ratio in fundraising per active cycle. Furthermore, since a number of government reports and news articles highlight the distinct expenditure patterns of scam PACs (Graham 2019; Janetsky 2018; Kleiner and Zubak-Skees 2019; Weintraub and Ravel 2016), I calculate PACs' average percentage of expenditures per active cycle across expense categories coded by the CRP.[3]

Table 5 reports summary statistics of PAC fudnraising and expenditure patterns. Three sets of findings emerge here. First, while overall scam PACs and legitimate PACs are comparable in the amounts of money they raised on average per active cycle, disparities emerge when comparing PACs within the same ideological leaning. In particular, average fundraising and expenditure by liberal scam PACs are roughly a half of those by liberal legitimate PACs. In contrast, conservative scam PACs raise and spend almost twice as much as those of conservative legitimate PACs.

[3]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 5: PAC-Level Summary Statistics

Statistic	Scam PACs (all)	Legitimate PACs (all)	Scam PACs (liberal)	Legitimate PACs (liberal)	Scam PACs (conservative)	Legitimate PACs (conservative)
No. PACs	84	1,871	9	774	69	911
Ave. total fundraising per cycle	$1,717,323	$1,661,383	$537,670	$1,196,692	$1,829,173	$1,031,937
Ave. total expenditure per cycle	$1,811,276	$1,783,882	$600,064	$1,258,548	$1,912,378	$1,128,497
Ave. itemization ratio of fundraising per cycle	0.381	0.838	0.436	0.813	0.389	0.849
% Expenditure on Contributions (CRP)	4.8%	23.0%	0.8%	27.4%	5.2%	19.8%
% Expenditure on Unclassifiable (CRP)	13.1%	7.8%	5.5%	7.6%	13.5%	7.3%
% Expenditure on Administrative (CRP)	16.6%	20.4%	16.8%	18.2%	16.1%	22.2%
% Expenditure on Non-Expenditures (CRP)	2.9%	4.1%	4.1%	3.2%	3.0%	4.8%
% Expenditure on Transfers (CRP)	0.3%	1.3%	0.3%	1.7%	0.3%	0.8%
% Expenditure on Strategy & Research (CRP)	3.7%	7.5%	9.6%	7.0%	3.4%	8.0%
% Expenditure on Campaign Expenses (CRP)	6.5%	2.7%	4.0%	3.2%	6.9%	2.0%
% Expenditure on Salaries (CRP)	2.9%	3.0%	16.0%	4.3%	1.9%	2.0%
% Expenditure on Fundraising (CRP)	33.7%	11.1%	13.1%	9.3%	34.8%	13.6%
% Expenditure on Media (CRP)	4.7%	7.8%	17.5%	6.9%	4.0%	7.8%

Second, regardless of ideological orientation, only about $38-44\%$ of all fundraising by scam PACs in a typical PAC-cycle came from itemized donors (i.e., those giving more than $200 to a recipient in a year), whereas the itemization ratio of legitimate PACs' fundraising is twice as high.

Third, scam PACs appear to pursue a markedly different strategy for campaign expenditures compared to legitimate PACs. Across the ideological spectrum, scam PACs channel a substantially smaller portion of its expenditures to campaign contributions to political candidates or other PACs ($0.8 - 5.2\%$) compared to that of legitimate PACs ($19.8 - 27.4\%$). This is consistent with the narrative that scam PACs seldom transfer electoral resources to political candidates or causes (Weintraub and Ravel 2016). In addition, within liberal-leaning PACs, scam PACs spend almost four times as much proportionally on salaries, which may be a sign of self-enrichment (Hunter et al. 2018). Within conservative-leaning PACs, scam PACs spend almost three times as much proportionally on fundraising, which appears be a main avenue of covert self-dealing activities among campaign consultants (Severns and Willis 2019).

5.2 *Itemized Donors*

I also examine a set of attributes of itemized donors that may help to distinguish scam PACs from legitimate PACs based on government reports and news articles. First, 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 many scam PACs appear to target ideological movements within parties (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 conservatism in a donor. Furthermore, some journalistic accounts of scam PACs suggest that their donors may be less habitual givers (Arnsdorf and Vogel 2016; Severns and Willis 2019). To measure donors' degrees of experiences with campaign giving, for each type of PACs examined I calculate their itemized donors' average number of distinct recipients (across different categories) as well as the average number of active cycles of giving.

Table 6: PAC Donor-Level Summary Statistics

Statistic	Donors of scam PACs (all)	Donors of non-scam PACs (all)	Donors of scam PACs (liberal)	Donors of non-scam PACs (liberal)	Donors of scam PACs (conservative)	Donors of non-scam PACs (conservative)
No. unique donors	212,503	3,436,961	79,940	1,385,472	132,327	211,749
% retirees	30.1%	9.9%	1.5%	13.1%	54.0%	40.2%
Ave. donor CFscore	0.42	-1.28	-1.65	-1.43	1.35	1.21
Ave. no. scam PACs given to	1.14	0.03	1.04	0.02	1.20	0.25
Ave. no. legitimate PACs given to	1.29	1.81	2.50	2.81	0.56	1.37
Ave. no. liberal scam PACs given to	0.39	0.02	1.04	0.02	0.00	0.00
Ave. no. liberal legitimate PACs given to	0.57	0.77	1.45	1.91	0.04	0.07
Ave. no. conservative scam PACs given to	0.75	0.02	0.00	0.00	1.20	0.25
Ave. no. conservative legitimate PACs given to	0.32	0.08	0.00	0.01	0.51	1.28
Ave. no. all recipients given to	19.78	7.33	34.36	11.24	22.83	12.46
Ave. no. active cycles	2.74	2.02	2.91	2.37	3.55	2.54

Table 6 compares these key measures of donor behavior across PAC categories. 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 6 need not generalize to all scam PAC donors (the vast majority of whom, as shown in Table 5, donate much less than the itemization threshold).

First, consistent with earlier results on the audience demographics of Facebook advertisements by PACs, notable age disparities emerge between scam PACs and legitimate PACs. Conservative-leaning PACs indeed exhibit the age patterns highlighted by the media (Graham 2019), where scam PACs appear to attract a higher fraction of self-reported retirees as itemized donors. Interestingly, this finding dovetails with existing research on how age appears to positively correlate with consumption of fake news and negatively correlated with digital literacy (Guess, Nagler, and Tucker 2019; Guess, and

Munger 2020). Nonetheless, the opposite appears to be true among liberal-leaning PACs, where itemized donors who have given to one or more liberal scam PACs are much less likely to report retiree status.

Second, itemized donors who give to scam PACs appear to be more ideologically extreme than those who give to legitimate PACs of comparable partisan or ideological orientation. Within liberal PACs, Itemized donors of scam PAC have much lower contributor CFscores on average, which suggest that they tend to be more left-leaning. In contrast, within conservative PACs, itemized donors of scam PACs appear to be more right-leaning as they have higher contributor CFscores on average.

Third, the *itemized* donors of scam PACs appear to be highly experienced at campaign giving. Across the ideological spectrum, itemized donors of scam PACs tend to have given to a lot more distinct recipients, and have made itemized contributions across more cycles (see the last two rows of Table 6). Moreover, while itemized donors of scam PACs tend to contribute less often to legitimate PACs of comparable ideological orientation, these donors do not appear to be unfamiliar with legitimate PACs. While these patterns appear to contradict journalistic claims that scam PAC donors are inexperienced at campaign giving (Arnsdorf and Vogel 2016; Severns and Willis 2019), they need not generalize to the vast majority of scam PAC donors who are unitemized (see Table 6).

5.3 *PAC treasurers and vendors*

Next, I analyze the personnel networks behind PACs' operations, specifically their treasurers and vendors. According to media reports, treasurers and vendors of scam PACs are often veterans in political consulting, and leverage their expertise as well as connections to found or serve scam PACs for personal financial gains (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 have worked for a greater number of PACs compared to their peers that only serve legitimate PACs.

To test this idea, I collect data on names and addresses of PAC treasurers and vendors as reported to the FEC. I 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 disambiguate identities using address information where appropriate. For each type of PACs examined, I calculate the average number of distinct PACs (across different categories) that their treasurers and vendors have worked for as well as the average number of active cycles of during which these treasurers and vendors worked for any PAC or campaign.

Table 7 displays summary statistics for PAC treasurers, and suggests that treasurers that have served on conservative-leaning scam PACs tend to have had worked for a larger number of PACs and for more election cycles in total compared to treasurers for conservative-leaning legitimate PACs. The opposite appears to be true among liberal-leaning PACs, although the distinction in work experience in campaigning between treasurers of liberal scam PACs versus liberal legitimate PACs is far less pronounced. In addition, Table 8 displays, for each category of PACs, the top 10 treasurers that have served the most number of such PACs. This tables mirrors patterns exhibited in the previous one. In particular, many of the top treasurers for conservative scam PACs, such as Dan Backer, Scott B. MacKenzie (i.e., the treasurer of the Conservative Majority Fund from my case study), and Paul Kilgore also are among the top treasurers for conservative legitimate PACs.

Table 7: PAC Treasurer-Level Summary Statistics

Statistic	Treasurers of scam PACs (all)	Treasurers of legitimate PACs (all)	Treasurers of scam PACs (liberal)	Treasurers of legitimate PACs (liberal)	Treasurers of scam PACs (conservative)	Treasurers of legitimate PACs (conservative)
No. unique treasurers	71	1,871	11	936	58	818
Ave. no. scam PACs served	1.51	0.02	1.09	0.00	1.57	0.05
Ave. no. legitimate PACs served	2.68	1.34	0.09	1.34	3.26	1.58
Ave. no. liberal scam PACs served	0.15	0.00	1.00	0.00	0.02	0.00
Ave. no. liberal legitimate PACs served	0.06	0.60	0.09	1.19	0.05	0.06
Ave. no. conservative scam PACs served	1.27	0.02	0.09	0.00	1.55	0.05
Ave. no. conservative legitimate PACs served	2.45	0.64	0.00	0.12	3.00	1.47
Ave no. any PACs served	17.75	4.30	2.27	3.75	21.22	6.13
Ave. no. active cycles	3.58	3.26	1.36	3.03	4.09	3.69

Table 8: Top Treasurers by PAC Type

Rank	Scam PACs (all)	Legitimate PACs (all)	Scam PACs (liberal)	Legitimate PACs (liberal)	Scam PACs (conservative)	Legitimate PACs (conservative)
1	Dan Backer	Josue Larose	Alexandra Rojas	Kinde Durkee	Dan Backer	Christopher Marston
2	Scott Mackenzie	Christopher Marston	Cary Peterson	Gregory Sanborn	Scott Mackenzie	Paul Kilgore
3	Kecia Pollock	Kinde Durkee	Francesca Lucia	Gary Crummitt	Robert Piaro	Nancy Watkins
4	Robert Piaro	Paul Kilgore	Grace Rogers	Jennifer May	Alexander Hornaday	David Satterfield
5	Alexander Hornaday	Nancy Watkins	Isra Allison	David Gould	Paul Kilgore	Lisa Lisker
6	Paul Kilgore	Cabell Hobbs	Keegan Goudiss	Rita Copeland	Zachary Bass	Cabell Hobbs
7	Zachary Bass	Douglas Edwards	Krystal Ball	Shawnda Deane	Ann Mattson	Dan Backer
8	Ann Mattson	David Satterfield	Michael Avenatti	Denise Lewis	David Satterfield	Scott Mackenzie
9	Christopher Marston	Lisa Lisker	Nathan Lerner	Judith Zamore	Kelly Lawler	Charles Gantt
10	David Satterfield	Dan Backer	Oliver Cappleman	Diane Evans	Paul Kutac	Barbara Bonfiglio

I also calculate an analogous set of summary statistics for PAC vendors shown in Table 9. Across ideological spectrum, vendors that have serviced scam PACs have worked for a lot more PACs and campaigns in total than those that have only serviced legitimate PACs. In addition, Table 10 displays, for each category of PACs, the top 10 vendors that have *exclusively* served the most number of such PACs.[4] Most of the vendors shown in Table 10 for scam PACs offer campaign consulting services (e.g., American Technology Services, Unified Data Services, Compliance Consultants), and echo prior accounts of how scam PACs' exclusive vendor networks help to obscure self-dealing activities among scam PAC officers (Lipton and Steinhauser 2015; Severns and Willis 2019).

Table 9: PAC Vendor-Level Summary Statistics

Statistic	Vendors of scam PACs (all)	Vendors of legitimate PACs (all)	Vendors of scam PACs (liberal)	Vendors of legitimate PACs (liberal)	Vendors of scam PACs (conservative)	Vendors of legitimate PACs (conservative)
No. unique vendors	3,383	49,742	367	28,302	3,045	17,878
Ave. no. scam PACs served	1.48	0.05	2.82	0.05	1.52	0.14
Ave. no. legitimate PACs served	4.29	1.50	18.40	1.71	4.32	2.13
Ave. no. liberal scam PACs served	0.14	0.01	1.26	0.01	0.05	0.01
Ave. no. liberal legitimate PACs served	1.83	0.78	10.47	1.37	1.65	0.48
Ave. no. conservative scam PACs served	1.31	0.05	1.43	0.04	1.45	0.12
Ave. no. conservative legitimate PACs served	2.28	0.57	7.10	0.29	2.49	1.59
Ave no. any PACs served	47.40	7.87	200.07	10.21	48.60	15.75
Ave. no. active cycles	2.46	1.89	2.61	1.91	2.50	2.21

[4] Regardless of whether we examine scam PACs or legitimate PACs, the most common vendors within category tend to be retail outlets such as hotels and airlines as well as mail vendors such as the USPS. As a result, I present Table 10 to highlight vendors that exclusively each category of PACs.

Table 10: Top Treasurers by PAC Type (Exclusive)

Rank	Scam PACs (all)	Legitimate PACs (all)	Scam PACs (liberal)	Legitimate PACs (liberal)	Scam PACs (conservative)	Legitimate PACs (conservative)
1	American Technology Services	Clark Hill PLC	Brand New Congress	Democratic Party Of California	National Capital Bank	Dinsmore Shohl
2	Unified Data Services	BBT Corp	Noble Christopher	House Majority PAC	Community Cares United	Wilson Perkins Allen Opinion Research
3	Compliance Consultants	Sonoma Restaurant	Progressive Rags	Global Strategy Group	Computerwild Inc	Williams Jensen
4	C Terry Raben	Squarespace Inc	Wendy S Wallberg Esq	Civis Analytics	EWH Small Business Accounting S.C.	Targetpoint Consulting
5	Catur Consulting	Harland Clarke Co	YouCanBook.me	Diablo Group	Fox, O'Neill & Shannon, S. C.	Advantage Direct
6	National Capital Bank	Majority Strategies	Zane Benefits	GBA Strategies	Hammen Michelle	Capital Cornered
7	TPFE Inc	Public Policy Polling	21c Hotels	National Democratic Club	Lifeline Services	Mckenna Long Aldridge
8	Community Cares United	Rightside Compliance	A2Z Convenience and Smoke Shop	PCMS LlC	Melissa Stetler	CC Advertising
9	Computerwild Inc	Wiley Rein LLP	Achievement Consulting	Capitol Compliance Assoc	Michelle Hammen	Mcintosh Co
10	EWH Small Business Accounting S.C.	Democratic Party Of California	Adroll Inc	Smart Final	Nielsen Merksamer	National Republican Senatorial Cmte

6 Supervised Machine Learning

The key features detailed in the previous section that distinguish scam PACs from legitimate PACs help to inform my construction of a supervised machine learning algorithm that systematically detects likely scam PACs. This section describes how I produce such an algorithm and analyzes its predictions.

6.1 *Sample selection*

For the purpose of model training, among the 84 scam PACs and 1,871 legitimate PACs that I identified using the procedure outlined in Section 3, I retain 78 scam PACs and drop the remaining 6 that lack data on their inferred ideological leanings (see Table 2). Among the remaining scam PACs, 9 are liberal leaning while 69 are conservative leaning. Similarly, I retain 1,505 of the original 1,871 legitimate PACs due to missing data. Among the remaining legitimate PACs, 823 are conservative leaning while 682 are liberal leaning. Since PACs' donor characteristics and personnel networks differ markedly by PACs' partisan or ideological orientation (Barber, Canes-Wrone, and Thrower 2017; Martin and Peskowitz 2015, 2018), it is unlikely that a supervised algorithm trained on conservative-leaning scam PACs will generate reliable predictions of liberal-leaning scam PACs and vice versa. As a result, I currently focus on estimating a supervised model that detects conservative-leaning scam PACs; the low number of liberal-leaning scam PACs makes it challenging to estimate a reliable model for them.

6.2 *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 legitimate PAC, as well as the probabilities that a given PAC falls into each of the two categories.

6.3 *Supervised algorithm*

The objective of this machine learning application is to distinguish (conservative-leaning) scam PACs from legitimate PACs based on their publicly observable attributes. To formally describe my supervised algorithm, let N_{train} be the set of PACs whose type (i.e., scam vs. legitimate 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 6.4). 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 $f(\cdot)$. By restricting my attention to detecting conservative-leaning scam PACs, it is reasonable to believe that $\hat{f}(\cdot)$ could identify conservative-leaning scam PACs not included in training data. If so, I can use $\hat{f}(\cdot)$ to predict whether any PAC in a test data set is more likely to be a conservative-leaning scam PAC or a legitimate 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 legitimate PAC.

6.4 *Feature selection*

I use a variety of model features to predict whether a given PAC is a scam PAC or a legitimate PAC. First, I include a number of PAC-level covariates based on findings shown in Table 5: 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 two summary statistics of itemized donors that have contributed to each PACs following conclusions drawn from Table 6: the share of itemized donors of a given PAC that self-reported as retirees, and the average 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 6. 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.

Fourth, 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 7 and 9, 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 linked to fewer than 12 PACs in the training set as well as treasurers or vendors linked to fewer than 2 PACs in the training set. While doing so reduces the number of model features, it considerably lowers the computational cost of model estimation. Moreover, model tuning results suggest that the marginal gain in predictive performance from including additional features is likely to be very small. This filtering process in total leaves $5,285$ unique donors, 57 treasurers, and 465 vendors for model training.

Last but not least, although PACs' communication materials via Facebook could also reveal their likelihood of being scam PACs, as shown in Section **??**, I exclude such materials from model training due to the presence of substantial missing data.

6.5 *Model Fitting*

I randomly selected 3/4 of the conservative-leaning scam PACs as well as 3/4 of the legitimate 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 stratified 10-fold cross-validation where each fold has an equal number of conservative-leaning scam PACs. The estimation procedure partitions the training set into my 10 specified subgroups and fits the model each time while holding one of the 10-sets out of sample.

For the purpose of model tuning, these cross-validation runs help me select the optimal value on the number of model features to be randomly sampled at each split during random forest estimation. The best value turns out to be 518, which corresponds to about 8.9% of the total number of predictors.

6.6 Estimation Results

Here I evaluate model performance by assessing predictions for the test data set. Since my data sample is heavily unbalanced (i.e., most PACs are classified as legitimate PACs), a natural concern is that an algorithm that predicts every PAC to be a legitimate 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 5 displays the receiver-operating-characteristic (ROC) curve for held-out predictions of scam PACs. The AUC in this case is 0.901. As a benchmark, if one naively predicts all PACs to be legitimate 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).

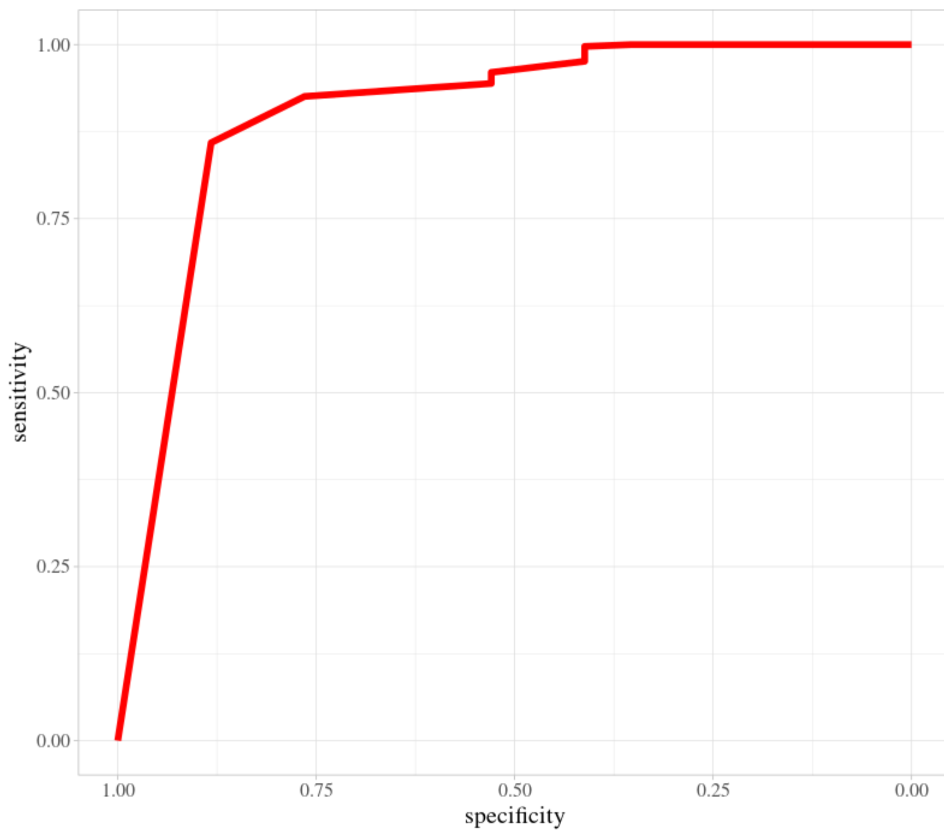

Figure 5: ROC Curve for Held-Out Predictions of Scam PACs

In addition, Table 11 shows the confusion matrix for predictions of scam PACs in the held-out sample based on my model. As this table shows, the model is able to distinguish scam PACs from legitimate PACs, albeit with errors. The false positive rate here is 0% (i.e., 0 out of 5 predicted scam PACs), and the false negative rate is 3.1% (i.e., 12 out of 388 predicted legitimate PACs). While at a glance Table 11 may suggest that the estimated model fails to identify a large share of the scam PACs in the test data set, auxiliary analysis suggests that these apparent false positives are mostly scam PACs that have received fewer allegations in the media. Within the test data, the predicted probability of a PAC being a scam PAC is positively correlated with the number of distinct news articles (if any) alleging said PAC as a scam PAC at 0.726. Insofar as the number of news mentions proxy for the credibility of such allegations, this high degree of correlation suggests that my algorithm may be able to identify bona fide scam PACs while limiting the risk of false alarms.

Table 11: Out-of-Sample Confusion Matrix

Prediction \ Reference	legitimate	scam
legitimate	376	12
scam	0	5

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 12 in descending order of predicted probabilities. In addition, Table 13 displays all scam PACs that are falsely labeled as legitimate PACs by my algorithm.

Table 12: List of All PACs Predicted As Scam PACs

PAC Name	Sample	Observed PAC Type	Predicted PAC Type	Predicted Probability of Being a Scam PAC	No. News Mentions
Us Veterans Assistance Foundation	training data	scam	scam	1	3
United American Veterans	training data	scam	scam	1	1
Americans for Law Enforcement	training data	scam	scam	1	7
National Campaign PAC	training data	scam	scam	1	3
Grassroots Awareness	training data	scam	scam	1	3
Standing By Veterans	training data	scam	scam	1	3
Firefighters and Emergency Responders Coalition	training data	scam	scam	0.952	3
Americans for Police and Trooper Safety	training data	scam	scam	0.952	2
Cops and Kids Together	training data	scam	scam	0.952	3

Our Country Deserves Better	training data	scam	scam	0.952	1
Action Coalition	training data	scam	scam	0.952	2
United Veterans Alliance Of America	training data	scam	scam	0.952	1
Voter Education	test data	scam	scam	0.952	3
Association for Emergency Responders and Firefighters	test data	scam	scam	0.952	3
U.S. Virgin Island Republican Party (VIGOP)	training data	scam	scam	0.905	8
Great America PAC	training data	scam	scam	0.905	5
Campaign To Defeat Barack Obama	training data	scam	scam	0.905	2
National Draft Ben Carson PAC	training data	scam	scam	0.905	2
Americans for The Cure Of Breast Cancer	test data	scam	scam	0.905	7
Stop Hillary PAC	training data	scam	scam	0.857	2
Heroes United	training data	scam	scam	0.857	3
Tea Party Patriots	training data	scam	scam	0.857	1
Conservative Majority Fund	training data	scam	scam	0.857	19
The Police Officers Support Association	training data	scam	scam	0.81	5
The Madison Project	training data	scam	scam	0.81	1
National Send Them Packing Committee	training data	scam	scam	0.81	1
American Coalition for Injured Veterans	training data	scam	scam	0.81	1
Put Vets First	training data	scam	scam	0.81	2
America Fighting Back	training data	scam	scam	0.81	1

Protect Our Future	training data	scam	scam	0.81	3
Republican Majority Campaign	training data	scam	scam	0.81	10
MAGA Coalition	training data	scam	scam	0.762	1
Black Republican	training data	scam	scam	0.762	1
Conservative Strikeforce	training data	scam	scam	0.762	20
Black Americans To Re-Elect The President	training data	scam	scam	0.762	2
Western Representation	training data	scam	scam	0.714	1
Breast Cancer Health Council	training data	scam	scam	0.714	1
Life and Liberty PAC	training data	scam	scam	0.714	5
Bampac	training data	scam	scam	0.714	2
Conservative Action Fund	training data	scam	scam	0.714	4
Restore American Freedom and Liberty	training data	scam	scam	0.714	4
Tea Party Leadership Fund	test data	scam	scam	0.714	4
Rightmarch.com	training data	scam	scam	0.667	3
Amish PAC	training data	scam	scam	0.667	1
Freedom's Defense Fund	training data	scam	scam	0.667	4
Republican Majority	training data	scam	scam	0.667	1
Tea Party Majority Fund	test data	scam	scam	0.667	10
Justice-PAC	training data	scam	scam	0.619	1
Stop Pelosi PAC	training data	scam	scam	0.619	1
Committee To Restore America's Greatness	training data	scam	scam	0.619	2
One Nation	training data	scam	scam	0.619	1
BLAKPAC	training data	scam	scam	0.571	1
Patriots for Trump	training data	scam	scam	0.571	1

Patriot Super PAC	training data	scam	scam	0.524	1
Conservative Freedom Fighters	training data	scam	scam	0.524	1

Table 13: List of All False Negatives in Predictions

PAC Name	Sample	Observed PAC Type	Predicted PAC Type	Predicted Probability of Being a Scam PAC	No. News Mentions
Patriots for Economic Freedom	test data	scam	legitimate	0.476	3
Keeping America Great	training data	scam	legitimate	0.429	1
Bold Conservatives (F.K.A. Draft Sherriff David Clarke for U.S. Senate)	test data	scam	legitimate	0.429	6
RallyPAC	training data	scam	legitimate	0.381	1
Tea Party forward	test data	scam	legitimate	0.19	1
Coalition Of Americans for Political Equality (CAPE)	test data	scam	legitimate	0.19	1
Go Big Go Bold	test data	scam	legitimate	0.095	1
Bikers for The President	test data	scam	legitimate	0.095	1
Constitutional Rights PAC	test data	scam	legitimate	0.095	1
Combat Veterans for Congress	test data	scam	legitimate	0.095	1
Draft Newt	test data	scam	legitimate	0.048	1
American Horizons	test data	scam	legitimate	0.048	2
Make America Great Again PAC	test data	scam	legitimate	0	1
Americans for Progressive Action USA	test data	scam	legitimate	0	1

6.7 *Feature Importance Results*

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 legitimate PACs, or that the information they provide is duplicated by that of other model features.

Table 14 displays the list of top 30 model features in terms of variable importance measured by mean decrease in Gini coefficients; it also includes the bivariate correlation between each feature and (conservative) scam PACs. This table shows that besides aggregate patterns of PACs and PAC donors that appear to distinguish scam PACs from legitimate PACs, PACs' linkages to individual donors, treasurers, and vendors improve scam PAC discernment as well, thereby further demonstrating the value of a supervised learning approach to detecting scam PACs (which allows for a large number of predictors) relative to classifying scam PACs based on a small number of PAC summary statistics. Moreover, Table 14 provides a validated measure of observable characteristics that differentiate scam PACs from legitimate PACs, and may serve as a useful set of heuristics for donors who wish to avoid scam PACs in campaign giving.

Table 14: Top 30 Model Features by Variable Importance

Index	Mean Decrease Gini	Feature Type	Feature Name	Corr. Scam PACs
1	8.135	Aggregate Donor Attribute	Ave. Itemization Ratio In Fundraising	-0.372
2	4.029	Individual Vendor	American Technology Services	0.334
3	3.872	Individual Vendor	Ignite Payments	0.313
4	3.804	Individual Vendor	Unified Data Services	0.334
5	2.196	Individual Vendor	GSI Inc	0.289
6	2.187	Aggregate Donor Attribute	Ave. Contributor CFscore	0.195
7	2.049	Individual Vendor	National Capital Bank	0.289
8	1.729	% Expenditure Category (CRP)	Non Expenditures	-0.002
9	1.688	Individual Vendor	Politicause LLC	0.239
10	1.539	Individual Vendor	Market Process Group	0.289
11	1.445	Individual Donor	Tracy, P J	0.303
12	1.409	% Expenditure Category (CRP)	Contributions	-0.109
13	1.398	Individual Vendor	TPFE Inc	0.264
14	1.286	Individual Donor	Nostrand, Gerald H	0.288
15	1.193	Individual Donor	Smith, Preston L	0.255
16	1.183	% Expenditure Category (CRP)	Campaign Expenses	0.114
17	1.141	Individual Donor	Smith, Jack	0.226
18	1.126	Individual Donor	Roberts, Dorothy B	0.26
19	1.077	Individual Vendor	Compliance Consultants	0.334
20	0.974	Individual Donor	Fox, Eleanor J	0.211
21	0.941	Individual Vendor	American Airlines	0.113
22	0.882	Individual Donor	Eller, James L	0.24
23	0.852	Individual Vendor	Politicallaw	0.178
24	0.814	Individual Donor	Marshall, John	0.26
25	0.775	% Expenditure Category (CRP)	Administrative	-0.032
26	0.767	% Expenditure Category (CRP)	Unclassifiable	0.035
27	0.747	Individual Donor	Sennett, David	0.195
28	0.74	Individual Donor	Lebewohl, Alice O	0.271
29	0.72	Individual Donor	Sprankle, Joseph F	0.22
30	0.718	Individual Donor	Siegel, Herbert	0.077

7 Discussion

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 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 (Akerlof 1973). 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 legitimate PACs in terms of solicitation strategies, fundraising and expenditure patterns, and donor and personnel networks. Based on these findings, I then 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 legitimate PACs at scale.

A number of data collection efforts and model training strategies may improve my existing descriptive analyses and supervised algorithm. First, the Federal Election Commission collects little information about PAC treasurers and vendors aside from their names, addresses, and records of payments they received from PACs in exchange for services performed. Moreover, no standardized identifiers for PAC treasurers or vendors are currently available. A data set that identifies unique treasurers and vendors over time, as well as tracking their webs of financial ties, could not only serve as additional model features for supervised predictions of scam PACs, but also help to identify instances in which PACs may be engaging in self-dealing activities. To that end, I hope to gather detailed information on treasurers and vendors' employment histories and biographic details using sources such as LinkedIn, registries of campaign consultants published by the American Association of Political Consultants and Campaigns and Elections, and registries of various lawyer associations.

Second, the Conservative Majority Fund case study suggests that scam PACs may be more likely to have records of non-compliance with existing campaign finance laws or other applicable laws. Consequently, it may be useful to collect information on complaints filed with the Federal Election Commission about potential PAC misconduct (Ravel 2015), audits and other enforcement actions that the Commission has taken with respect to specific PACs (Butler 2015), and investigations undertaken by the Federal Bureau of Investigation and the Department of Justice (Stueve 2019, 2020). Again, such information could both provide useful model features for supervised machine learning and help to identify unambiguous scam PACs among all PACs that have received such allegations in media reports.

Third, the aforementioned data collection efforts can also allow me narrow down my sample of "legitimate PACs", which currently consist of all qualifying non-connected PACs that have yet to receive any allegations of being scam PACs in the news, to a subset of bona fide legitimate PACs (i.e., helping to reduce instances of scam PACs mistakenly labeled as legitimate PACs in my training data).

Fourth, for my analyses of communication strategies exhibited by scam PACs versus legitimate PACs, I would like to expand both the set of legitimate PACs included in the existing analyses of Facebook data, and incorporate other data sources on communication materials such as emails (Mathur et al. 2020) and samples of phone transcripts, mailers, and other means of offline solicitation (Severns and Willis 2019). Beyond expanding my data collection in these fronts, I hope to provide more validation results of text sentiment and topic modeling shown in Section 4, and explore the use of manipulation tactics and "dark patterns" (Mathur et al. 2020).

Last but not least, I would like to explore alternative models for supervised machine learning, such as dropout models (Srivastava et al. 2014), that may produce superior out-of-sample performance for rare-event detection.

Ultimately, I hope to use findings from this project to springboard field experiments, in partnership with government agencies or campaign organizations interested in combating scam PACs, that test the effectiveness of different information interventions for increasing donor discernment of scam PACs and changing subsequent donation patterns. One type of information intervention could be to simply alert donors about the existence of scam PACs in general or a list of verified scam PACs, which may be constructed using predictions from my supervised algorithm for scam PAC detection. The success of Charity Navigator and other non-profit rating agencies provides grounds for optimism that such interventions may be able to prevent campaign donors from falling victim to scam PACs, and redirect would-be scam PAC donors to supporting bona fide candidate campaigns or legitimate PACs that actually advance donors' objectives (Gordon, Knock, and Neely 2009; Yoruk 2016). However, political science research on misinformation casts doubt on the effectiveness of similar strategies, where making people aware of such problems may breed mistrust and reduce political participation overall (Ternovski, Kalla, and Aronow 2021). Another form of information provision could be to equip donors with a list of tips for spotting potential scam PACs (e.g., low itemization ratio in campaign fundraising, ties to specific treasurers or vendors that have served scam PACs), which may be deduced from my descriptive findings as well as the feature importance results of my supervised algorithm. There is encouraging evidence on similar types of information treatments in helping the public discern fake news, although whether better discernment translates into meaningful changes in political behavior remains an open question (Guess et al. 2020).

Results from such experiments could shed light on the usefulness of different data-informed approaches to increasing donors' ability to distinguish scam PACs from legitimate PACs, enhancing PACs' accountability to campaign donors, and restoring trust in the fundraising marketplace. Moreover, the field experiments I proposed may serve as an indirect test on the extent to which individual donors' campaign contributions are consumption-driven (Ansolabehere, de Figueiredo, and Snyder 2003; Barber, Canes-Wrone, and Thrower 2017). Specifically, if a series of well-powered field experiments consistently find that no amount of form of information relating to scam PACs can change donors' propensity to give to scam PACs versus legitimate PACs, one may interpret such null results as evidence that individual campaign contributions are driven by pure consumption values (e.g., donating to a PAC that purports to share a donor's political beliefs through inflammatory public rhetoric and solicitation tactics) rather than strategic considerations (e.g., donating to PACs that engage in forms of campaign spending that are productive to affecting election outcomes or issue advocacy). Such findings, if true, would have important implications for the communication and expenditure strategies that all candidates and PACs pursue since individual donors are by far the largest and growing source of campaign funds for federal candidates in the United States (Barber, Canes-Wrone, and Thrower 2017).

References

Ansolabehere, George. 1973. "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism." *Quarterly Journal of Economics* 84(3): 488–500.

Altman, Alex and Michael Scherer. 2014. "Conservatives Inc." *Time Magazine*, March 13. https://time.com/23001/conservatives-inc/ (March 23, 2021).

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.

Butler, David. 2015. *Reference: Amended April Quarterly Rreport (01/01/2014 - 03/31/2014), Received 10/15/2014*. https://docquery.fec.gov/pdf/896/15330083896/15330083896.pdf (March 24, 2021).

Charity Navigator. 2021. *Charity Navigator's Methodology*. https://www.charitynavigator.org/index.cfm?bay=content.view&cpid=5593#rating (March 25, 2021).

Charity Navigator. 2021. *Financial Score Conversions and Tables*. https://www.charitynavigator.org/index.cfm?bay=content.view&cpid=48 (March 25, 2021).

Center for Responsive Politics. 2021. *Total Outside Spending by Election Cycle, Excluding Party Committees*. https://www.opensecrets.org/outsidespending/cycle_tots.php (March 25, 2021).

CrowdTangle Team. 2020.[List ID numbers: 1513676; 1520325]

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

Federal Bureau of Investigation. 2020. *FBI Warns Voters About Election Crimes Ahead of the November 2020 Election*. https://www.fbi.gov/news/pressrel/press-releases/fbi-warns-voters-about-election-crimes-ahead-of-the-november-2020-election (March 24, 2021).

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.

Fowler, Erika Franklin, Michael M. Franz, Gregory J. Martin, Zachary Peskowitz, and Travis N. Ridout. 2020. "Political Advertising Online and Offline." *American Political Science Review* 115(1): 130–149.

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

Geraghty, Jim. 2019. "The Right's Grifter Problem."*National Review*, June 3. https://www.nationalreview.com/the-morning-jolt/the-real-problem-conservatism-faces-today/ (March 23, 2021).

Gordon, Teresa P., Cathryn L. Knock, and Daniel G. Neely. 2009. "The Role of Rating Agencies in the Market for Charitable Donations: An Empirical Test." *Journal of Accounting and Public Policy* 28: 469–484.

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).

Groseclose, Timothy and Jeffrey Milyo. 2005. "A Measure of Media Bias." *Quarterly Journal of Economics* 120(4): 1191–1237.

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*.

Guess, Andrew, Jonathan Nagler, and Joshua Tucker. 2019. "Less Than You Think: Prevalence and Predictors of Fake News Dissemination on Facebook." *Science Advances*. 5(1): 1–8.

Guess, Andrew and Kevin Munger. 2020. "Digital Literacy and Online Political Behavior." Working paper. https://osf.io/3ncmk/

Guess, Andrew M., Michael Lerner, Benjamin Lyons, Jacob M. Montgomery, Brendan Nyhan, Jason Reifler, and Neelanjan Sircarh. 2020. "A Digital Media Literacy Intervention Increases Discernment Between Mainstream and False News in the United States and India." *Proceedings of the National Academy of Sciences of the United States of America*. 117(27): 15536–15545.

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.

Lewis, Matt. 2015. "The 'Conservative' PACs Trolling for Your Money."*Wall Street Journal*, May 7. https://www.wsj.com/articles/the-conservative-pacs-trolling-for-your-money-1431040712 (March 23, 2021).

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.

Arunesh, Mathur, Carsten Schwemmer, Brandon M. Stewart, Angelina Wang, Maia Hamin, and Arvind Narayanan. 2020. "Manipulative Tactics Are the Norm in Political Emails." Working paper. https://electionemails2020.org/assets/manipulative-political-emails-working-paper.pdf

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.

Ravel, Ann M. 2015. "Stopping Scam PACs From Ripping Off Donors."*Roll Call*, July 13. https://www.rollcall.com/2015/07/13/stopping-scam-pacs-from-ripping-off-donors-commentary/ (March 24, 2021).

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).

Roberts, Margaret E., Brandon M. Stewart, and Dustin Tingley. 2019. "stm: An R Package for Structural Topic Models." *Journal of Statistical Software* 91(2): 1–40.

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).

Srivastava, Nitish, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. "Dropout: A Simple Way to Prevent Neural Networks fromOverfitting." *Journal of Machine Learning Research* 45(1):5–32.

Stueve, Joshua. 2019. *Treasurer of Political Action Committees Pleads Guilty to Filing False Reports with Federal Election Commission*. https://www.justice.gov/usao-edva/pr/treasurer-political-action-committees-pleads-guilty-filing-false-reports-federal (March 24, 2021).

Stueve, Joshua. 2020. *Political Consultant Sentenced for Fraud Involving Scam PACs*. https://www.justice.gov/usao-edva/pr/political-consultant-sentenced-fraud-involving-scam-pacs (March 24, 2021).

Ternovski, John, Joshua Kalla, and Peter Aronow. 2021. "Deepfake Warnings for Political Videos Increase Disbelief but Do Not Improve Discernment: Evidence from Two Experiments." Working paper. https://osf.io/dta97/

Tucker, Will. 2015. "Controversial PAC Conservative Strikeforce faces FEC questions."*OpenSecrets*, June 24. https://www.opensecrets.org/news/2015/06/controversial-pac-conservative-strikeforce-faces-fec-questions/ (March 24, 2021).

Tucker, Joshua A., Andrew Guess, Pablo Barberá, Cristian Vaccari, Alexandra Siegel, Sergey Sanovich, Denis Stukal, and Brendan Nyhan. 2018. "Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature." The Hewlett Foundation. https://www.hewlett.org/wp-content/uploads/2018/03/Social-Media-Political-Polarization-and-Political-Disinformation-Literature-Review.pdf

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).

Yoruk, Baris K. 2016. "Charity Ratings." *Journal of Economics and Management Strategy* 25(1): 195–219.