Institute for
New Economic Thinking

Working Papers
ineteconomics.org/workingpapers

Move Fast and Break Everything: Crypto, Democrats and Deregulation

**Thomas Ferguson, Jie Chen,
Matthias Lalisse and Paul Jorgensen**†

Working Paper No. 245

January 12th, 2026

Abstract

In November, 2022, the giant cryptocurrency exchange FTX filed for bankruptcy. The financial fallout from that event, including two bank failures, made crypto politically radioactive. Yet less than three years later, crypto, like Donald Trump himself, staged a triumphant Second Coming, as the President signed into law the so-called "GENIUS Act" – short for Guiding and Establishing National Innovation for U.S. Stablecoins. Analysts have traced the industry's phoenix-like resurrection, showing how key crypto billionaires and companies aligned with Trump early in the 2024 campaign, transforming him from a skeptic into a political champion. But the flip side of the story is much less discussed: how

† Thomas Ferguson is Professor Emeritus at the University of Massachusetts, Boston; Director of Research at the Institute for New Economic Thinking; and serves on the Academic Advisory Board of Better Markets. Jie Chen is University Statistician at the University of Massachusetts, Boston. Matthias Lalisse received his Ph.D. in Cognitive Science (Neural Networks) from Johns Hopkins University and works in AI and Machine Learning. Paul Jorgensen is Associate Professor, University of Texas Rio Grande Valley.

support for crypto has grown among Democrats. This paper analyzes voting by House Democrats on the GENIUS and Clarity Acts, in the context of the campaign for sweeping financial deregulation mounted by both crypto and traditional finance. The implications of a growing race to the bottom in financial regulation and emerging challenges in cybersecurity receive attention.

https://doi.org/10.36687/inetwp245

JEL codes: E42, G28, E58, K23, L51, P16

Keywords: crypto, financial deregulation, central bank digital currency, Donald Trump, money, regulation, campaign contributions

In November, 2022, the giant cryptocurrency exchange FTX filed for bankruptcy. The shocking collapse triggered runs on other crypto firms, forcing several to close, along with two major banks, while crestfallen venture capital and sovereign wealth funds suffered millions of dollars in forced write-downs.[1] So-called stablecoin funds, like Tether, were also hit, but survived.

Crypto looked not just battered, but on the brink of financial nuclear winter.

Yet less than three years later, crypto, like Donald Trump himself, staged a triumphant Second Coming. On July 18, 2025, President Trump signed into law the so-called "GENIUS Act" -- short for Guiding and Establishing National Innovation for U.S. Stablecoins. Along with the President, industry leaders and their allies rattled off a litany of benefits they claimed would flow from the legislation –from cheaper payments to a stronger worldwide demand for dollars.

Analysts in both the major media and smaller outlets have traced the industry's phoenix-like resurrection, showing how key crypto billionaires and companies aligned with Trump early in the 2024 campaign, transforming him from a skeptic into a political champion.[2] The President and his family embraced these new allies, invested heavily with some, and profited giddily from regulatory momentum that helped vault the President 118 spots on the Forbes 400 list of richest Americans.

The President's unwavering support, combined with the firehose of political money that industry stalwarts showered on Republican legislators, leaves little mystery about how crypto won over the Republican Party.

But crypto's conquest of the Republicans is only one part of its ascent. Less discussed is the story's flip side: how crypto is advancing among Democrats.

News coverage of Sam Bankman Fried and FTX suggesting that the firm's political strategy focused lopsidedly on Democrats was always off base; such claims rested on incomplete and superficial tabulations of the firm's overall political contributions (Ferguson, Jorgensen, and Chen, 2022). But the bankruptcy discouraged more subtle and detailed postmortems.

In the meantime, leading Democrats, including Biden, accepted some crypto contributions, but the President and his key regulators – especially then-Securities and Exchange Commission Chair Gary Gensler – remained highly skeptical. Top Federal Reserve officials, like nearly all leaders of major banks at the time, shared their reserve, which crypto's nuclear winter did nothing to thaw. So in 2024, when legislation friendly to crypto came to votes in the U.S. House of Representatives, Republicans waxed enthusiastic, but most Democrats followed the administration's lead in rejecting proposed legislation.[3]

But in 2025, something changed: Fully 102 House Democrats voted for the GENIUS Act, and 78 – more than double the number expected –joined with the Republicans to pass a companion

“Clarity Act” pushed by crypto advocates to define key regulatory issues beyond stablecoins. (Unlike GENIUS, the Clarity Act stalled in the Senate; it has yet to pass.[4])

Given the far-reaching consequences of crypto's ascendence (including fast developing markets in crypto derivatives), this shift among the Democrats deserves more scrutiny. The House, with more than 200 Democrats voting, offers far greater statistical traction than the much smaller Senate, where sample sizes constrain any meaningful analysis.

We examined voting by House Democrats for both the GENIUS Act and the Clarity Act. This paper's “Statistical Appendix” sets out our model in full; here we explain our findings in non-technical terms. Surprisingly, what did not explain these votes proved as revealing as what did. The Trump-Harris vote differences in districts did not matter. Neither did lawmakers' margins of victory, their ages, or state-specific factors that might hide other influences. Instead, crypto support correlates closely with wealth, ideology, financial contributions, and one other district feature discussed below.

In the case of the GENIUS Act legislators from districts with more high-income households were more likely to vote for pro-crypto bills: for every 1% increase in households earning over $200,000, the odds of a yes vote rose by 7.4%. More conservative Democrats -- based on several standard measures of Congressional ideology -- were also more likely to vote yes, with one or two notable outliers among progressives.[5]

The total amount of money legislators received from crypto interests helped predict their votes. For every $1,000 increase in total contributions, the odds of voting for the GENIUS bill increase by 0.2%. Since contributions often amounted to many thousands (sometimes even hundreds of thousands) this effect cannot be neglected.

Data Quality and How We Handled It

As much as we appreciate the valuable efforts of some public platforms to track campaign finance, we have repeatedly found that no published totals can be relied upon for rigorous analysis, even when they are recycled in well-known academic studies.[6] We have detailed our reservations at length before; our separate Data Appendix explains what we did for this study. That Data Appendix includes a detailed discussion of the algorithm we developed to analyze the contribution data along with a codebook linked to a file available on the website of the Institute for New Economic Thinking.

The key point: we took a deliberately conservative approach to counting crypto money. We did not count all donations from all venture capitalists – or, a fortiori, from all of “Red Tech” -- even though many such donors likely support crypto and some may hold peripheral investments in it. The numbers we report concern firms and investors clearly tied to crypto, including some of the

Forbes 400 richest Americans, along with a handful of active political champions, lobbyists, and supporting professionals.

History Repeats

Our results echo our previous study of Congress and the systematic weakening of Dodd-Frank financial reforms. That research exploited an exceptionally favorable case that helped to rule out potential confounding variables. Divisions within the industry and active opposition from other sectors were minuscule. Control of the presidency or major regulatory agencies did not turn over in the period under study. By design, the research could focus on Democratic legislators who initially voted *for* Dodd-Frank but later voted to weaken it. We could isolate more clearly how money influenced behavior (Ferguson, Jorgensen, and Chen, 2020).[7]

The pattern was striking: each additional $100,000 in contributions significantly raised the odds a legislator would vote to weaken financial regulation by 13.9%. Some members received substantially more -- and voted accordingly. This wasn't speculation; it was a clear, predictable pattern of behavior change correlated with financial incentives.[8]

The fallout was real. Under Trump (and Jerome Powell), the Federal Reserve chipped away at the new, tighter Dodd-Frank banking regulations in a series of small but meaningful steps. When Biden's team attempted a reset, particularly around Basel III capital requirements, their effort petered out as Trump reemerged as a champion of deregulation in the run up to the 2024 rematch. By January 2025, top regulators at the Bank of England and the Bank for International Settlements were openly voicing fears that the United States could set off a race to the bottom in global financial regulation.

A stark sign of the prevailing winds came from news that UBS, the giant Swiss bank, was seriously considering moving to the U.S. to escape stricter oversight.

This is larger context in which the struggle between crypto and the mega banks is best appreciated, because adding crypto to this mix creates distinctive new dangers.

Setting aside bitcoin and meme coins – which are all essentially gambling, as the recent roller coaster swings in the value of so many illustrate again – the real policy battle centers on stablecoins. Proponents claim these represent genuine technological innovation, a "better mousetrap" for payments and remittances.

We are skeptical.

Technology Claims Don't Hold Up

The blockchain technology underlying stablecoins remains relatively slow – despite ongoing development, it cannot handle high transaction volumes. It is also clunky, often expensively so. Only some stablecoins are designed to traverse across blockchains and the transits, like old time railway interchanges across the borders of European countries, come with enforced stops (tolls). Many are plainly intended to function like walled gardens, trapping their users in specially cultivated ecosystems that makes exits costly. Switching between stablecoins involves fees and charges, not to mention exiting the universe altogether.[9] And stablecoins are not – at least yet – federally insured, so that runs, when they happen, are likely to be lightning fast, rather like those in the major bank collapses in 2023.[10]

Even within a single denomination, decentralized blockchain technology may not really be suitable for a global payment system. Cryptocurrencies are premised on the idea that secure payment and monetary integrity can be derived from decentralized networks of computers and token-holders that certify transactions without requiring any central issuer, ledger-keeper, or enforcer. Some analysts have recently suggested on theoretical grounds that this underlying security mechanism has poor scaling characteristics in comparison to conventional rule of law, since the latter enjoys major economies of scale and an ability to freely calibrate the magnitude of punishment to deter attacks (Budish, 2025). By contrast, crypto relies on scale—the amount of resources (computation, staked capital) devoted to processing payments—which relates to the cost of commanding a significant proportion of the network to insert malicious blocks. But since the rewards for a successful attack swell with the network's value and usage (more and bigger wallets to pick), so would the resources needed to deter attacks. These costs are paid one way or another (new tollbooths, capital parked to implement payment processing, or risk, etc.). As more users enlist, more cash and circuits must be burned to enforce trust on the blockchain. With the costs of electric power skyrocketing as Large Language Models proliferate, such misgivings can only increase.

At a more fundamental level, if blockchain technology could be made as swift and efficient as stablecoin advocates envision, central banks could simply implement one themselves and provide the service directly to citizens at very low cost. Such a universal system would allow the resulting economies of scale to benefit the public rather than private oligopolies and facilitate innovation in the rest of the economy by reducing costs of operations for everyone else.

Right now, for example, the Brazilian central bank runs a hugely successful instant payment system that makes transfers at no cost to individuals and at much lower rates than private credit cards for businesses. It does not run on blockchain, but on a database owned and managed by the central bank, which also superintends its cybersecurity. According to the International

Monetary Fund, encrypted payments in the system settle in seconds, compared to two days for debit cards and 28 days for credit cards.

The lesson for regulators and the public is to beware hype about "21st century systems": all payment systems transmit at roughly the speed of light even if the amount of information flowing through them varies. Their economically important differences lie in information checks, security protocols, regulatory supervision and, crucially, how many intermediaries take toll along the way. Plus, of course, who gets to hold the means of payment and for how long.[11]

Tellingly, all the interest groups in the struggle over the GENIUS and Clarity Acts strongly oppose central bank digital currencies. Neither traditional banks nor stablecoin issuers want competition from government-provided digital payment systems.

This alignment between the rivals speaks volumes, and it is striking how each side's arguments in favor of its payment scheme evolve over time. Keeping payments data out of the hands of government was long a universal rallying cry, but it is now so obviously specious it is heard less and less. Advocates just talk past how simple it is for data brokers to buy and resell people's "private" data. Indeed, lobbying battles are now erupting over whether banks can be forced to share client data with crypto firms.

The Remittance Question

Our study yielded one unexpected result: districts with higher percentages of Hispanic constituents were more likely to see their legislators vote for pro-crypto legislation. For every 1% increase in a district's percent Hispanic, the odds that its representative voted for the GENIUS Act increase by 3.1%. Since percentages vary substantially, the effect is sometimes sizeable. Initially this result puzzled us, but a review of claims by advocates in and out of Congress reveals heavy emphasis on how crypto might lower the costs of international money transfers.[12] These matter greatly for immigrant communities sending money to families abroad; by far the largest flows involve Mexico (Congressional Research Service, 2023).

Whether these cost savings materialize in practice is hard to say. The collision between traditional financial systems and crypto is creating complex, evolving market dynamics. Traditional bank remittance fees remain high, though fees and charges for getting in and out of blockchains are sometimes sizeable. Banks—which once complained bitterly about "know your customer" regulations—now sometimes quietly point to them as an advantage for customers.[13]

The underlying problem is that major banks have long since abandoned poorer customers. A 2023 FDIC study found that almost a fifth of the U.S. population has little or no access to banking services.[14] The situation is reminiscent of the well-known food deserts in many large cities: a recent Federal Reserve Bank of Atlanta study showed that highly consolidated grocery chains often consider doing business in low-income areas not worth the effort, seeing the potential rewards as too small.

Independent studies of the costs of remittances suggest that competition in sending money abroad has failed, just like it has in credit cards. A report by the Bank for International Settlements and the World Bank Group pointed out the “thin profit margins for basic account providers [of banking services]. Some respondents noted that there is little economic incentive for private sector parties to provide these accounts.”[15]

This market failure has given crypto firms an opening they shouldn't have. The solution isn't necessarily crypto deregulation—it's ensuring financial institutions serve all communities adequately with controls on crypto that are equivalent to best practice in banking.

The Dark Side of Crypto

The remittance question is inseparable from crypto’s role in criminal activity. For all the glowing praise of blockchain transparency, many crucial parts of crypto are anything but transparent. It is easy to disguise the ownership of crypto wallets, for example.[16] Ransomware attackers almost always demand payment in crypto, typically Bitcoin. Corporate executives have described watching ransom payments vanish into the blockchain, bouncing between wallets and exchanges before ending up in casinos and splitting into countless untraceable fragments.

One of the clearest depictions of the scale of the problem is Jeff White's *Rinsed*. Crypto ATMs are overwhelmingly concentrated in poor neighborhoods, facilitating various forms of money laundering. Most allow users to buy crypto, but not usually to get out of it. Criminals recruit individuals to conduct transactions by providing small amounts of money on bank cards—a common laundering technique that operates both domestically and internationally.[17]

The philanthropic activities abroad touted by some stablecoin proponents deserve scrutiny through this lens. The blockchain isn't just a technology—it's an ecosystem of intermediaries including crypto wallets, brokers, and dealers, many operating in highly concentrated networks that allow operators to move money in and out of customers' pockets with minimal oversight.

Big Money and How Lawmaking Really Works

The GENIUS Act appeared to prohibit interest payments on stablecoins (Wired, 2025a). The provision was a major factor persuading traditional banking groups not to move strongly against the legislation. That prohibition has already broken down. One major stablecoin issuer is already advertising 10 percent returns for "loaning out" stablecoins through affiliated but technically separate entities. All sides are appealing to regulators and to Congress to do something.

The new legislation in theory requires stablecoin issuers to function as “narrow banks” – they are allowed to hold only very short term, (hopefully) highly liquid assets that can be sold on demand against any run on their stablecoins. The supervisory structure to enforce this is complicated and is in fact a work in progress. It relies on public attestations by firm executives and audits by accounting firms as early screening devices as well as supervision by different regulators depending on the size of institutions. But with regulators under budgetary and political pressure, how this will work in practice remains to be seen. Reminding many observers of the bad old days of pre-Civil War wildcat banking, the system appears highly vulnerable. The recent downgrade of Tether by S&P Global Ratings is a fire bell ringing in the night.

The situation is paradigmatic for the problems of a money-driven political system. When big money flows freely, crucial details get worked out long after a law has passed, in heavily technical discussions far away from public consideration, under pressure to conserve regulators' time and resources, including some who are transiting from one revolving door to another.

Cyber Storm on the Horizon

There is more, alas. Both finance and crypto now face challenging new problems: above all, cybersecurity.

Here is where the regulatory race to the bottom meets galloping technology. The new administration drastically curtailed many important regulatory bodies as soon as it came to power, by cutting their budgets or simply sweeping them away. It dismissed the Justice Department's National Currency Enforcement Team. It eliminated the Corporate Transparency Act that Treasury Secretary Janet Yellen had championed to control shell companies and improve financial transparency and neutered the Consumer Financial Protection Agency. Under Trump, the SEC has also pursued a unique, hands off policy toward crypto. These aren't random policy adjustments – they represent a systematic dismantling of the regulatory infrastructure vital to preventing financial crime, just as technological changes pose dramatic new types of threats.

Alarmingly, the latest statement of the new administration's National Security strategy does not even mention cybersecurity as a priority.[18]

The omission is consequential. Under Biden, the Cybersecurity and Infrastructure Security Agency (CISA) struggled to convince companies to take cybersecurity seriously. Big budget cuts and the government shutdown have left it reeling under a new acting director who failed a polygraph test. Even at its old strength, the task was daunting: Competing firms are very reluctant to acknowledge failures, as the chilling example of Solar Winds illustrates.[19] Companies that do tackle security risks can find themselves losing business to more feckless competitors, as has been shown vividly in discussions of insurers who underpriced risks before recent California wildfires.

Private insurance markets cannot solve such problems because, as Daniel Schwarcz and Josephine Wolff (2025) have documented, insurers cannot adequately price cyber risk based on actual security incidents and measures taken—in practice they simply charge by firm size or even sector.

The only effective approach is mandatory standards—telling companies a minimum set of safeguards they must put in place. Without such requirements, vulnerabilities will only worsen.

The advent of quantum computing poses this challenge at a wholly new level. At the moment quantum hackers have cracked only very elementary encryption systems. Bitcoin and other systems use much stronger encryption, but it is plain that the combination of artificial intelligence, quantum computing, and autonomous AI agents is creating giant new risks at precisely the moment government regulators appear to be retrenching dramatically. Some months ago, Sam Altman of Open AI warned that the security systems many financial houses rely upon were now easily penetrated by hackers (Associated Press, 2025). A Federal Reserve governor allowed that this was a question that the Fed could perhaps study in collaboration with the tech giants. Since then, however, misgivings about federal government's interest in cybersecurity have only grown. As one headline summarized the situation at the end of the year, "Fears Mount That US Federal Cybersecurity Is Stagnating—or Worse." (Wired, 2025b).

Conclusion: Two Giant Systems Colliding

The US is now in the midst of a collision between two giant financial systems—traditional banking and crypto—with neither adequately regulated to serve the public interest and both lobbying for still more deregulation. Political money sloshes everywhere in the system, shaping outcomes just as it weakened Dodd-Frank and enabled crypto's political resurrection after FTX.

The crypto debate is not really about technology or innovation – it's about power, money, and whether democratic institutions can assert the public interest over private profits.

The situation might be summarized thus: Current crypto systems combine low cost and low security. They may evolve into something more useful, but right now, as a group and, often, individually, they present serious risks that money-driven politics is systematically obscuring – as for example in the emerging crypto-derivatives market. Meanwhile, the regulatory infrastructure needed to protect the public –from ransomware, money laundering, cyberattacks, and financial instability—is neglected or being actively dismantled, as federal enforcement pulls back.

Our research reveals a final significant trend, which we have space here only to mention, not discuss in full. Not just the Republicans, but the Democratic Party is now awash in crypto money. Industry donations run very high indeed to the National Committee and congressional

Democratic Leaders, such as Senate Minority Leader Charles Schumer, House Minority Leader Hakeem Jeffries, or the House Minority Whip, Katherine Clark, and some individual legislators. Depending on the election cycle, the sums sometimes run in the millions, not tens or even hundreds of thousands. Our check of the Internal Revenue Service's roster of 527 funds, which are not counted in the Federal Election Commission tabulations, also reveals crypto cash flowing abundantly to many state Democratic party organizations. These are not reported to the Federal Election Commission, but to the IRS, where they are ignored by the press. It is clear that these massive flows will be a silent backdrop to the debates now raging in the Party over its positions in the 2026 Congressional elections and beyond.

Statistical Appendix

This section sets out our statistical model of voting on both the GENIUS and Clarity Acts discussed in our main text.

We start by defining the dependent variable. Let $Y_i = 1$ if the representative voted in favor of the GENIUS legislation and $Y_i = 0$ if the representative voted against; and we let $p_i = P(Y_i = 1)$. This is a binary, not a continuous variable, so we employ logistic regression rather than ordinary least squares.

House districts are numerous and often geographically contiguous, which raises potential concerns about spatial autocorrelation (Ferguson, Jorgensen, and Chen, 2022). We thus first test for spatial dependence. Because the dependent variable is binary, spatial autocorrelation was assessed using Bonferroni-adjusted joint count tests (Cliff & Ord 1981; Anselin 1988). These tests compare the observed number to their expected values under spatial randomness. The results indicated no statistically significant spatial clustering, suggesting that spatial autocorrelation is not a concern for this study.

Logistic regressions rely on a log-odds transformation. Their coefficients are commonly interpreted by reference to odds ratios, that is, how a unit increase in the predictor changes the odds of the outcome under study. In this case, the outcome is a vote in favor of the GENIUS bill. Our table also shows results for the Clarity Bill, defined similarly to the vote on GENIUS. We did not assume that that an equation similar that for GENIUS would necessarily work for Clarity; but that is what our tests indicate.[20]

Note that these are less intuitive than the more familiar R^2 measures in OLS. We calculate one common pseudo R^2, Nagelkerke's.

The equation we estimate is:

$$ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1\text{\%Hispanic} + \beta_2\text{\%Income2000} + \beta_3\text{Nokken1} + \beta_4\text{Nokken2} + \beta_5\text{TotalMoney222426}$$

Nokken 1 and 2 refer to the Nokken-Poole (or DW-Nominate) estimates of representatives' ideologies in two dimensions; those are calculated with reference to one Congress rather than several.[21]

We have always been cautious about using these measures. Recently many other scholars have also expressed skepticism. See, for example (Morris, 2023) or (Desilver 2022) The first dimension is typically regarded as a member's position on a left/right scale and is most widely discussed in the press. The meaning of the second dimension is increasingly debated. But we found the later useful in this case, so we have retained it.

Total Money is calculated for the three election cycles of 2022, 2024, and the current (2026), with a cut off at the end of July, 2025 for the that cycle. The download date was August 19.

	GENIUS			Clarity		
Predictors	Estimate (β)	*Odds Ratio* (e^{β})	*p-value*	Estimate (β)	*Odds Ratio* (e^{β})	*p-value*
Hispanic (%)	0.028	1.028	0.015	0.029	1.030	0.014
Household over $200,000 (%)	0.064	1.066	0.007	0.072	1.074	0.002
Nokken1	.099	1.105	<0.001	.098	1.103	<0.001
Nokken2	.046	1.047	<0.001	.053	1.055	<0.001
TotalMon222426 ($1,000)	0.002	1.002	0.008	0.001	1.001	0.051
Observations		212			212	
Pseudo-R^2 (Nagelkerke)		.592			0.593	
Area Under the ROC Curve (AUC)		.898			.903	
Hosmer_Lemeshow Test		$\chi^2(8) = 5.012$, p = .756			$\chi^2(8) = 14.045$, p = .081	

Note that: Nokken1 = Nokken-Poole Dimension 1 ×100

Nokken2 = Nokken-Poole Dimension 2 ×100

Interpretation

For the GENIUS bill, the estimated coefficient for Hispanic is 0.028 and the odds ratio is 1.028, indicating that for every 1% increase in a district's percent Hispanic, the odds of voting for the bill increase by 2.8%. Thus for every 10% increase in the percent Hispanic, the odds of voting for the bill increased by 31.8%, ($1.028^{10} \approx 1.318$).

The estimated coefficient for the percent of Households over $200,000 is 0.064 and the odds ratio is 1.066, indicating that for every 1% increase in households making more than $200,000, the odds of voting for the bill increase by 6.6%. The odds of voting in favor of the bill are about 1.895 times greater for 10% increase in Hourseholds_200000ormore, ($1.066^{10} \approx 1.895$.**)**

The estimated coefficient for Nokken-Poole Dimension 1 ×100 is 0.099 and the odds ratio is 1.105, indicating that for every 100 units increase in Nokken-Poole_dim1 the odds of being GENIUS = 1 increase by 10.5%. Because of the way the index is scored, this means that more conservative Democrats are more likely to vote in favor of the bill, though some notable exceptions exist.

The estimated coefficient for Nokken-Poole 2 ×100 is 0.046 and the odds ratio is 1.047, indicating that for every 100 units increase in Nokken-Poole_dim2 the odds of voting for the bill increase by 4.7%.

The estimated coefficient for Money is 0.002 and the odds ratio is 1.002, indicating that for every $1,000 increase in money the odds of voting in favor of the GENIUS bill increase by 0.2%. The distribution of contributions is skewed, with many running into tens or even hundreds of thousands of dollars.

The estimated coefficient for **Money** is 0.002, corresponding to an odds ratio of 1.002. This implies that for every $1,000 increase in contributions, the odds of voting in favor of the GENIUS bill increase by approximately 0.2%. Note that the distribution of contributions is highly skewed.[22] 21.6% of all Democrats received no contributions, 39% less than five thousands; 55.4% less than 10 thousands, and 80% below $100,000.

For the **Clarity**, the estimated coefficient for Hispanic is 0.029 and the odds ratio is 1.030, indicating that for every 1% increase in the percent Hispanic, the odds of voting for the bill increase by 3.0%. For every 10% increase in the percent Hispanic, the odds of voting for the bill increased by 34.4%, ($1.030^{10} \approx 1.344$**.)**

The estimated coefficient for the percent of Households_200000 is 0.072 and the odds ratio is 1.074, indicating that for every 1% increase in Households over $200,000 the odds of voting for the bill increase by 7.4%. The odds of voting in favor of the bill are about 2.042 times greater for 10% increase in Hourseholds_200000ormore, ($1.074^{10} \approx 2.042$.**.)**

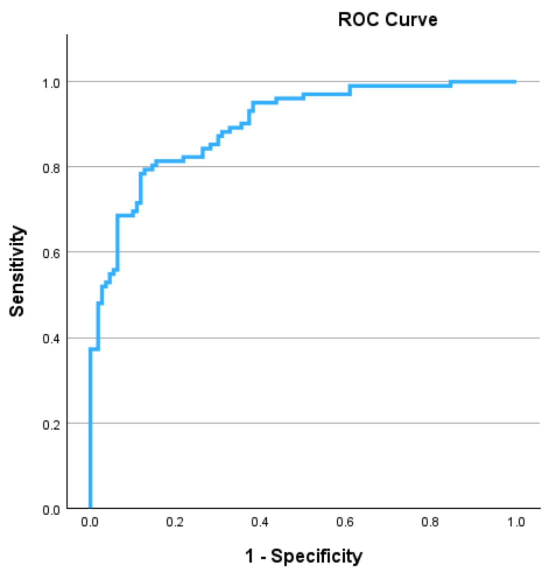
The estimated coefficient for Nokken-Poole Dimension 1 ×100 is 0.098 and the odds ratio is 1.103, indicating that for every 100 units increase in Nokken-Poole_dim1 the odds of being GENIUS = 1 increase by 10.3%. Because of the way this is scored, this means that more conservative Democrats are, again, more likely to vote in favor of the bill.

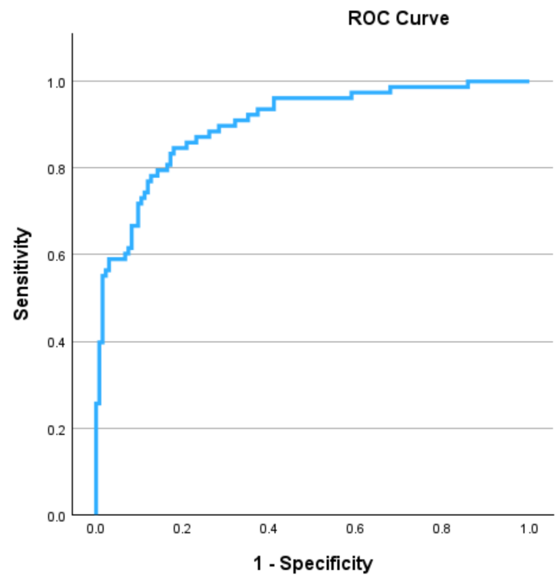
The estimated coefficient for Nokken-Poole 2 ×100 is 0.053 and the odds ratio is 1.055, indicating that for every 100 units increase in Nokken-Poole_dim2 the odds of voting for the bill increase by 5.5%.

The estimated coefficient for Money is 0.001 and the odds ratio is 1.001, indicating that for every $1,000 increase in money the odds of being GENIUS = 1 increase by 0.1%.

A common way to assess the predictive performance of a logistic regression is the Area Under the Receiver Operating Characteristic (ROC), often abbreviated as AUC. The AUC ranges from 0.5, indicating no better than chance prediction, to 1.0, indicating perfect discrimination. The estimated AUC for the GENIUS model is .898, suggesting excellent predictive performance.

GENIUS	Clarity
Area Under the ROC Curve Test Result Variable(s): PF Area: .898	**Area Under the ROC Curve** Test Result Variable(s): PF Area: .903





We of course tested for multicollinearity. Multicollinearity diagnostics indicated no concern in either model (maximum VIF = 1.091; condition index = 8.5).

Data Appendix and Codebook for the Campaign Contribution Data

Matthias Lalisse, Paul Jorgensen

and Thomas Ferguson

Building datasets that truly represent investors and institutions active in particular arenas is not easy, as we have explained many times (Ferguson, Jorgensen, and Chen, 2013, 2023). The multiplicity of data sources does nothing to help. The Federal Election Commission is the principal source of data on contributions to federal elections, but the Internal Revenue Service is a second that is only occasionally consulted. These two sources format their data in completely different ways and can only with difficulty be integrated together. State authorities offer data of varying quality for state and some local elections.

This study concerns contributions to Democrats in Congress, so the FEC is the mother lode, though we also consulted data from the IRS, as mentioned in our main text. Here we explain how we constructed the dataset and, crucially, the use we made of machine learning techniques to process it. At the end of this appendix, we append the Codebook to our dataset of contributions, which is separately posted on the website of the Institute for New Economic Thinking.

Campaign reporting is typically organized in two-year cycles denoted by year of the actual election itself along with the year before. Political action committees (PACs) are major sources of contributions. They are easier to identify and study than individuals, so many studies focus just on them – a practice we have many times cautioned against.

For this paper, we employed a two-pronged strategy to identify cryptocurrency donors in FEC and IRS filings in the three election cycles of 2022, 2024, and 2025 (the first year of the 2026 cycle). We began by assembling a list of cryptocurrency PACs. Most of these are associated with one or a handful of particular companies, though people contribute through them who are not necessarily company employees at all.

First, starting from a list of cryptocurrency PACs, we identified both individual and institutional donors then traced their contributions to candidates or committees. We assembled the PAC list itself from multiple sources, including earlier studies we published (Ferguson, Jorgensen, and Chen, 2023); other industry-wide reports (Claypool 2024; followthecrypto.org; the Center for Responsive Politics); known affiliations with crypto business firms (Coinbase, GMI, Web3Forward, Payward Ventures), along with PACs we identified ourselves as associated with digital assets, blockchain, and crypto.

Our round-up included the gargantuan industry Super PACs that have been widely reported on. These have been prominent in all three of the election cycles we study in this paper and the degree of concentration among them has been very high indeed.[23] The full list is provided here.[24]

Individuals and organizations contributing to such committees report in their FEC filings. Using these core nodes (the crypto pacs) as starting points, we trace these donors' contributions to other committees and candidates, eventually finding an additional $209 million spent by the industry outside of the more visible channel of crypto committees registered with the FEC. Along with their super PACs, these 316 donors to the core set of crypto committees accounted for about 89% of the money from the crypto industry in 2021-2025 ($685 million).

Along with backpropagating crypto affiliations through the PAC list, we ran string searches for large cryptocurrency firms for self-reported names and employers (specific firms like Coinbase, Pantera Capital, Riot Platforms, Gemini… as well as generic industry terms: Blockchain, DAO, Bitcoin, and others). Employer and organization name-level searches were supplemented with comprehensive string searches for individuals named in various Forbes crypto billionaire lists. Search terms were designed as flexible regular expressions with manual filters on false positives (hits on the search terms were checked for correctness on each retrieved donor). With the string searches, we found an additional 988 cryptocurrency donors contributing an additional $78 million. In sum, we found $765 million of crypto contributions flowing into US elections in the three election cycles that we analyze.

But, as we have observed many times, the real problems only begin once the preliminary list of names and donors are assembled. To predict congressional votes, one generally wants the dollar amount that ended up in the campaign accounts controlled directly by the candidate. This flow of funds approach is difficult to implement; it is a principal reason why efforts to study political money resist simple inventorying. Committees that receive contributions from crypto or any other individual, company, or industry, often transact among themselves but do not add new money into the election cycle. These inter-committee transactions present challenges to the analysis of campaign finance data.

The FEC dataset contains many duplicate transactions. These duplicates stem from committee reporting practices. In some cases, the donor committee and the recipient committee both report the same transaction. In other cases, candidates enter fundraising arrangements with conduits like ActBlue or WinRed. These conduits report processing the contribution in addition to the FEC reporting that same contribution as the donor contributing to the candidate directly. Another such conduit, albeit more complicated, is the Joint Fundraising Committee. Candidates who join these committees do so with other candidates and/or other committees (such as state parties). The initial contribution to the joint fundraising committee is then divided among the committee's constituent members.

Cash contributions split up among committees should in principle be siftable and add up to the original lump sum of money. In practice, though, imperfections in federal accounts and reporting systems can create the political money equivalent of the famous Schrödinger Cat in quantum physics. Depending on how you trace, the chain of transfers can sometimes just peter out, as someone or something in the chain fails to report or records the transaction in a form different from what it really was. The errors can combine in luxuriously kaleidoscopic ways, not least because the subdivision formulas along the daisy chain are generally not publicly divulged (ultimately, the candidate is supposed to report having received them) and because relevant data may occasionally be reported in financial reports filed long after the original transactions. We thus take care to note the filing dates of reports we rely on.

In our analysis, we only include crypto contributions given directly to the candidate, or crypto money spent explicitly in support of or in opposition to a candidate. Crypto spending against candidates is subtracted from their crypto total. Regarding those Joint Fundraising Committees, we conservatively only include the contributions that make their way through those committees to be received by the candidate-controlled committees directly. In most cases, this decision includes money contributed to the candidate's principal campaign committee and/or the candidate's leadership committee.

Accurately labeling contributions as crypto is no easy task. FEC data are self-reported, with many donors (whether duplicitously or not), reporting employment statuses that do not reflect their actual employer or economic group. “Retired,” “Not Employed,” or “Information Requested” (but not obtained by the recipient committee) are common entries in the self-reported employers of many large contributors. Many studies have failed to correct for this reporting bias when investigating the effects of campaign contributions on Congress (e.g. trying to predict votes on Medicare legislation by lumping together all self-reported “Retired” donors—including those spending many times the average social security check or more on political contributions. The actual dependents of that program, though more numerous when measured as people rather than dollars, barely appear when aggregated together with the more affluent donors from their cohort).[25]

Sometimes, misreported employers are easy to spot. An investment venture or a workforce commanded is not always “employer” (from a certain point of view), and many large donors hold positions on boards of directors (whether of firms or charities), some while being nominally “retired”. Generally, for large donors, at least one transaction in the heap will point to the individual’s main economic affiliation. Therefore, a first de-noising step required in studies of campaign finance is to merge donor records across variations in how they report themselves to each committee. This task, known as deduplication or record linkage, is a core flywheel in our methodology.

But it goes without saying that, as many Large Language Models now caution users of their output, AI can indeed make mistakes. Manual corrections of data remain an important part of our processing, though less with each development cycle. Our advice to anyone discussing individual political figures would be to check yet again.[26]

Our algorithm and training data for donor merging has been tuned iteratively, and continues to be improved as we gather more training data and identify useful features.[27] The Federal Election Campaign Act of 1971 (as amended) requires federal political committees to undertake “best efforts" collect identifying information from each donor, including name, mailing address, employer and occupation, with these descriptors publicly reportable whenever a donor contributes more than $200 to a committee within any calendar year (Congressional Research Service, 2025).

"Best efforts" is a far from an exact term. They are generally sufficient to get donor-reported descriptors, but these are notoriously noisy and often highly misleading, particularly among large donors who may have multiple employment affiliations.

Other cases of misreporting are more mysterious. In the 2024-25, we find about 1,200 contributions from individual donors with verifiable connections to the cryptocurrency industry who either reported being unemployed, self-employed, or retired, with no further description (or who did not report any employer). This is 7% of our total list of transactions—a consequential proportion, accounting for $1.5 million in transactions, $846,000 of which went to political candidates. Not inconsiderable flows of money are thus missed if employer affiliations are approached at the level of the transaction rather than the donor.

It is possible to penetrate the veil for most (but not all) such instances by first throwing a lasso around each donor: inducing clusters of transactions from the same entity based on partial hits cross-referenced across fields. As long as one of the grouped records contains the true industry affiliation, the donor is retrieved in company-level searches. The clustering methodology we use for this task is based on the dedupe software package, which implements the record linkage framework developed by Bilenko (2006) to cluster multi-field text records by (1) first grouping plausible matches into parsimonious data blocks (2) that are then assigned co-membership probabilities using similarity functions learned from training data.

Off the shelf, the dedupe package is designed to initialize with few training data points, making use of regularization (both explicit and through model simplicity: l2-regularized logistic regression) to assign feature weightings and cluster membership probabilities with training samples numbering in the tens. The core methodology runs swiftly even on datasets consisting of millions of potential duplicates, the key innovation (Fellegi and Sunter, 1969; Bilenko 2006) being that matches are sparse relative to the number of possible comparison pairs, with only a faint subset of these comparisons requiring direct labeling using the computationally costly similarity function.[28]

We leave this backbone untouched. When accuracy is paramount, however, the similarity function itself leaves much to be desired. Therefore, in addition to having assembled a very large training set of labeled pairs consisting of more than 30,000 examples (ratio 1:2 of match to distinct), we employ an extensively customized version of dedupe that allows free selection of a similarity function kernel. This is further augmented with extensive feature engineering that improves separation of the target classes, particularly when models are nonlinear. Regarding the choice of similarity kernel itself, we have found ensembles of decision trees (random forests) effective in generating graded similarity scores, i.e. scikit-learn's `predict_proba` method for the

RandomForestClassifier class (proportion of leaf-node training points matching the consensus class, averaged across the ensemble). In addition to their well-known power on high-dimensional problems, RFs can allow more concise feature codings than is desirable in link-linear specifications (e.g., logistic regression), leading to smaller models, and empirically they outperform logistic regression on our problem (see Ablation results below, Table A.2).[29]

Regarding the feature set, our linkage model uses a combination of standard string match metrics on the donor identifiers along with a large collection of hand-engineered features aimed at improving true match rates while numerically encoding heuristic rules for licensing merges (e.g., "require strong evidence from employer or zip code for a commonplace name like 'Jim Jones'"), enabling good class separation for match and distinct. A summary of supplementary features is found in Table A.1.

Table A.1: Features of the link model

Feature name	Num feats	Description
String match		String distance (affine gap metric) between matching fields. We match on the base fields from FEC records (NAME, EMPLOYER, OCCUPATION, CITY, STATE, and ZIP_CODE (truncated to 5 digits), as well as a battery of features derived from the base fields by parsing and text normalization. We expand person names to GivenName, MiddleName, and LastName for individuals, and BaseName, CorpOrgType (corporation, llc, gmbd, pac) for organizations. Derived fields use custom normalization scripts as well as the probablepeople parser[1] to disentangle name parts.
String length	2	Min-max (symmetric) encoding of string length for each item of the pair. Two features.
State	1	Strict equality (e.g. CA is not closer to VA than TX).

1 https://probablepeople.readthedocs.io/en/latest/ We have experimented with generative parsing using LLMs with greater accuracy/customizability, particularly for organization names, for which no adequate parser exists when high-level denoising is required (e.g., "ABC Investments" and "BCE Investments" have high string similarity in toto, but separating the (all-important) within-industry discriminant of the brand, low string similarity. Extant off-the-shelf parsers are not configurable for extracting the relevant segments—an obvious use-case for generative AI. While LLMs struggle with the demands of high-throughput applications like these, as with many tasks in NLP, they also obviate much of the classical pipeline.

Nicknames	1	“Bill” and “Will” are similar, but “Elisabeth” and “Lizzy” are not. We expand each base name to its list of nicknames, and return 1 if the two sets overlap, 0 otherwise.
Probable gender	1	Match of the probable gender based on the GivenName. These are mere guesses based on self-reported gender identity: name-to-gender conditionals estimated from a large dataset of person names packaged through the gender-guesser utility.[2] The feature is important for distinguishing near-neighbors (e.g. “Alex”, “Alexa”) where string distance is confounded.
Person-name frequency	2	How common is a name? We decompose name frequency into two features, one for the first, and another for the last. E.g., “Mike” and “Johnson” will have many hits, “Haim” and “Saban” has few. Algorithm is more stringent in other fields on common name pairs.
K-Street penumbra	1	Enumerated zip codes for the Washington DC area including parts of Virginia and Maryland. We observed a high density of lobbyists reporting from residential locations in the penumbra in addition to workplaces located within-capital. Political action committees tend to cross-list addresses in organization headquarters/candidate home state and also in DC. Feature allows matching across these common variations conditioning on entity type.
Union locals	2	Unions marked for whether a local or a national (trigger, phrases like ‘local’ or ‘district council’). If a local, the identifier for the local is extracted via regex pattern: ‘^.*\blocal\b\W*((n(o?)(\.?)|num((\.|ber)?)))\W*([1-9][0-9]*)\b\.*$’, extracting the numeric part, and checking for equality. In trainset labeling, locals are kept separate from other locals and from the national.
Labor	1	Prominent unions are enumerated—UAW, IATSE, etc.—with verbal expansions (“United Auto Workers”). Enumeration is merged to a large disjunctive regular expression. Centered sum-of-match-any coding—score(name1,name2)=any_match(name1) + any_match(name2)-1.

2 https://github.com/lead-ratings/gender-guesser

Lexical categorical: Individuals	4	Indicators for common reported employer/occupation types (retired, self-employed, unemployed, homemaker), which are often confounding for large donors (see main text). Phrase-detection uses regular expressions applied to the concatenation of Employer and Occupation strings.
Lexical categorical: Organizations	~10	Regular expressions for detection of certain indicators of organization type in NAME field for organizations, such as "political action committee"/"PAC", "incorporated"/"inc."/"company"/"co.", "senate"/ "congress"/"president", e.g. to differentiate "Timothy Scott for Senate/President". Coding for indicators is centered sum-of-match-any.
Zip5-depth	1	Depth of the digit-by-digit match for a zip-code pair, truncated at the first mismatch. Proxy for geographic distance.
Distance (geographic)	1	Geodesic distance between zip-level GPS coordinates. Metric: Haversine, using precompiled nominatim coordinates packaged through pgeocode.[3]
Zip5-level population, population density	6	Large cities contain more individuals, with therefore higher risk of name collisions. To calibrate match stringency, we use year-to-year census estimates of population in the donor's zip code, combined with land area for the tract. Allows for greater stringency on other features for donors with co-incident location in high-density areas. Max-mean representation of each base feature (raw population for the zip code, raw land area, and computed population density for the tract). Mean-max representation allows the values for both records to be represented without breaking symmetry.
Vectors	3	Lawyer and Attorney are dissimilar as strings, but similar as concepts. Rather than enumerating all such synonyms, or attempting to do the same for all firms (Google, Alphabet), we encode all occupation and employer strings into vectors and compare them using the cosine similarity. We use the Mixedbread AI encoder (transformer base, mxbai-embed-large-v1/1024d). Fields encoded are: Employer, Occupation, and (for organization-type donors) Name. We use context-grounded vectorization prompts (e.g., "Encode the following employer for

3 https://github.com/symerio/pgeocode

		comparison with features of similar employers:") to improve separation. A sub-problem of high importance is ensuring high throughput for string comparison given the invocations of the vectors that are quadratic in the size of the blocks. Therefore, these are precompiled for each field, taking dot products at runtime.
The Jr/Sr problem	N/A	Deterministic postprocessing routine, not a feature as such. By design, generational suffixes append to identical or near-identical individual names (e.g., Bill Gallop Jr/Sr). Even if generational suffixes are separated out, these too have high string overlap (II is very similar to III or IV). More problematically, generational suffixes are often not explicitly marked, with the unmarked instances interpolating between the marked ones in fuzzy clustering, leading to record merges that include incongruent suffixes. Worst of all, the Roman numbering system may coexist with Jr/Sr marking: Bill Gallop II may enter his name as Bill Gallop Sr, with Bill Gallop III being the Junior. Such clusters are methodically detected and split into marked (one for each suffix found) and unmarked. If both present, Roman and Sr/Jr as: Sr to the seniormost Roman partition, Jr to their immediate progeny in the line. This forms the marked partitions. For unmarked instances, normalized similarity from other fields determines final assignment.
Duplicative forking	N/A	Daniel Michael Day-Lewis (of *Lincoln* and *There Will Be Blood*) has two surnames. Should the actor make political contributions, he might reasonably do so as Daniel Lewis, which entries are not blocked with either Daniel Day or the full name, the surnames lacking a common prefix. To improve blocking rates in similar actual instances (of which there are many), we fork donor entries with multiple hyphenated or unhyphenated surnames, then merge forked records to the clusters for the root entry in postprocessing.

Ablation of algorithm components

We have made serious efforts to evaluate our pipeline, which processes FEC bulk releases in four-year increments. Because of the huge size of the transaction tables (recent elections contain some 70-80 million rows by the end of the election cycle), algorithmic tagging is a necessity, with manual cleanup whenever errors do occur in the donor identification step. To assess the pipeline itself, we have conducted an ablation study, isolating the contribution of two macro-components of the underlying algorithm: the classification algorithm (random forests), and the supplemental features. Two baselines are considered: the full-feature model with logistic regression as the scoring model (LogReg), and random forests with comparison of only the base fields (Name, Employer, Occupation, City, State, Zip) using only string distance comparators (RF-String).[4] Each ablated model was trained on the same 30,000-pair training set to induce clusters on the full 2024 FEC transaction tables, combined with early data for the 2026 cycle (up to August 2025).

We report precision and recall for each specification (full model and ablated) on both all crypto donors (individuals, PACs, and companies), and on the individual subset, where the two baselines are more competitive. Table A.2 displays the results, with cluster-wise precision and recall (averaged across manually verified donor clusters) for full and ablated models. Our extant production build (many-feature RF) ranks best along both performance dimensions (precision and recall). Results were poor for both baseline models on organizational donors, with substantial entrainment of unrelated organizations (low precision) in organization clusters drawn by both models. For the logistic model, this defect is in part attributable to the fact that indicator feature encoding in the full-feature baseline (denoted lexical-categorical in Data Appendix Table A.1) is non-monotonic in the match level (note that indicator features are important distinguishing committees that are similar by both string and semantic measures, e.g., “Menendez for Congress” and “Mendez for Senate”). Feature design could therefore be further optimized for linear weighting, but as an exercise to guide system design, the results are enough to convince us that many-feature random forest configuration is most performant, with the difference particularly pronounced in precision (number of transactions falsely attributed to a single donor). The current error rates appear more than tolerable, especially when backed by manual care, and we believe they represent a considerable advance over earlier efforts.

4 The RFString baseline includes as well a categorical meta-variable for entity type.

Data Appendix Table A.2: Ablation of algorithm components

Model	Precision	Recall	Precision (IND only)	Recall (IND only)
LogReg	.761	.987	.801	.994
RF-String	.908	.975	.964	.983
RF-Full	.989	.999	.988	.999

Sources for other variables used in our study.

1. GENIUS Act Vote: https://clerk.house.gov/Votes/2025200
2. Clarity Act Votes: https://clerk.house.gov/Votes/2025199

3.Vote Margins for Harris and Trump Vote by Congressional Districts used in 2024.

https://docs.google.com/spreadsheets/d/1ng1i_Dm_RMDnEvauH44pgE6JCUsapcuu8F2pCfeLWFo/edit?gid=1491069057#gid=1491069057

3. Congressional District Voting Results for 2024 – From Leip, Atlas of US Presidential Elections, US Representative General - County & Congressional District Level Vote Data.
4. Nokken-Poole estimates for House Members. Note these are keyed for that Congress rather than across time. Nokken, Timothy P., and Keith T. Poole. 2021. "Congressional Voting Data." *Voteview*. https://voteview.com/articles/data_help_members.
5. Household Median Income, Household Mean Income, % of Households over $200,000: % Hispanic or Latino -- ACS 2024 Estimates for Congressional Districts. Other demographic groups tested for came from ACS 2023 extimates.

Crypto Congress Dataset: Codebook

The data come from an August 19, 2025 download of the FEC bulk transactions dataset (the FEC's itoth and itcont) that we have intensively combed through. Users will need to bear in mind the discussion above of the significance of such dates. Candidates are linked to their committees, including leadership PACs and other committee types that the FEC does not track (such as single-candidate super PACs). We include all transactions that targeted a Democratic member of the 119th Congress. Since this paper analyzes the Democrats, they are the subject of our most detailed processing and checks, as outlined in our Data Appendix. Our preliminary assessment of spending on Republican candidates and party committees suggests that crypto heavily favored Republicans in a ratio of about 2 to 1. Note that while we analyzed crypto as a bloc, the data are supplemented with per-donor IDs that we used to group all transactions by a donor entity (PAC, company, or individual), even if the entries for industry affiliation were unmarked or incomplete in the corresponding FEC report.

There is no shortage of million-dollar-plus contributions that made their way through to candidates. Per-donor amounts are calculated from the nonduplicate rows (see DeduplicatedContrib below for details). Amounts flowing from companies and individuals to congressional campaigns, however, often understate the flow attributable to such donors, since super PACs served as the conduit for the bulk of their political spending. Note that the amounts differ systematically between these different types of contributions, with contributions that flow through the Citizens United ecosystem being limited only by the supply of donors. Two subtypes of political money are represented in the data: ads run by crypto-affiliated super PACS (66%) and contributions to candidate-affiliated political committees, some of which are also super PACs, but most subject to conventional campaign finance limits. We differentiate between independent spending for and against a candidate (FEC transaction types 24A). While negative ads were run by crypto super PACs against multiple House members, only one emerged from the fray to claim a seat in the 119th Congress (Andrea Salinas). While ads supporting a candidate or payments to their political spending accounts are straightforward to interpret, ads against complicate analyzing targets. This is an important methodological concern, especially when one needs to simultaneously analyze multiple candidates running in the same race (see Ferguson, Jorgensen and Chen 2019 for extensive discussion), where the candidate- and even the party-coding of a primary-stage contribution can be ambiguous. Even in studies of seated-member voting (where only the only contributions that matter are those that affected the winner), expenditures against can be relevant, since contributions to them can signal the donor disfavors the ultimate winner—although, with less visibility on contributions to opponents' campaigns, interpreting the contribution for a vote might be arguable. For 24A contributions, close readings of the ads themselves are sometimes sufficient to decrypt which opponent is implicitly endorsed. Other times, we can lean on firmer evidence, either from donors' public endorsements or from their other contributions. In the single case of a candidate targeted by 24As in our dataset, no complicated arguments are necessary: the super PAC at issue (Protect Our Future, associated with FTX) also spent millions on ads for her Democratic opponent Carrick Flynn, the likely intended beneficiary of the ad. Tracing the sprawl of political money in its many forms and variegated effects is without doubt a valuable enterprise and cases of this sort are relatively rare and difficult to parse. For the purposes of the data released with our results, the important technical note is that we treated such cases as unidimensional and directional, so the sparse 24A expenditures (affecting Oregon's 6th district in 2022) were projected on the same scale as the rest of the money but with a minus sign.

Below we provide descriptions of each included column to ease use:

Column name	**DataType**	**Description**
DonorID	String	ID for the donor linking contributions by same entity. For IDs with format Cxxxxxxxx (x numeric), the donor is a PAC.
DonorName	String	Name of the donor (individual, company, or PAC).
DonorType	String	$\in$ {PAC, IND, ORG}. Type of entity the donor is.
Employer	String	Employer if IND donor (as it appears in the FEC filing).
Occupation	String	Occupation if IND donor (as it appears in the FEC filing).
City	String	City reported by the donor (as it appears in the FEC filing).
State	String	State reported by the donor (as it appears in the FEC filing).
ZipCode	String	Zip Code reported by the donor (as it appears in the FEC filing).
DeduplicatedContrib	Indicator {0,1}	Indicator for whether the contribution represents a duplicate transaction: 1 if the contribution is new, 0 otherwise. Including transactions with 0 in this column leads to double counting. We correct for the following types of duplicates: • Amended reports • Transactions to intermediate committees that simply transfer to other political spenders (ActBlue, WinRed, joint fundraising committees, and similar) • Doubly-reported PAC contributions (PACs that contribute to other PACs—the contribution appears in two reports)
TransactionType	String	FEC transaction type. See https://www.fec.gov/campaign-finance-data/transaction-type-code-descriptions/ for descriptions of each code.
TransactionAmt	Integer	Size of the contribution in nominal US dollars. We manipulate the dollar amount in two instances: refunds, and independent expenditures against the candidate (see intro text above).

TransactionDate	DateString	Date of the transaction in lexicographic format: YYYY-MM-DD
RecipientPAC_ID	String	If the contribution was received by a PAC (excludes independent expenditures), the FEC identifier (Cxxxxxxxx, x numeric) for the PAC.
RecipientPAC_Name	String	Name of the recipient if a PAC.
FEC_CandID	String	FEC identifier (Hxaaxxxxx, x numeric and aa a state code) for a candidate running for a given office. May be missing for committees whose affiliations are not tracked by the FEC.
CandID	String	Identifier for the candidate that is constant across districts, offices, and election cycles.
BioguideID	String	Unique entity identifier assigned to each incoming congressional class by the Biographical Directory of the United States Congress.
CandidateName	String	Name of the candidate.
CandidateOffice	String	Office run for by the candidate. The format is office level {H, S, P}, followed by state if not president, followed by the district if House, separated by underscores.
RecipCommitteeDsgn	String	If the recipient is a committee, the designation of that committee. P denotes a campaign committee, D a leadership PAC, and J a joint fundraising committee. U-type (unauthorized) committees include single-candidate super PACs among other committee types, and their affiliations are not tracked by the FEC.
RecipientParty	String	Party of the candidate targeted by the contribution. Note that some Democrats double-brand with the Democratic Farmer-Labor party (DFL) or some other.
SUB_ID	String	Unique numeric row identifier for an entry in the FEC's bulk datasets. These occasionally blink in and out of existence as the FEC updates its database, but they provide the only sure means to map from our dataset to the original source. We prefix the FEC's internal number with s to coerce read-in as a string, since software reading the entries as an integer will encounter overflow and collapse records.

References

American Banking Association Banking Journal. 2025. "House Passes Bills on Stablecoins, Digital assets, CBDCs," July 17, 2025; https://bankingjournal.aba.com/2025/07/house-passes-bills-on-stablecoins-digital-assets-cbdcs/

Anselin, L. 1988. *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer. https://doi.org/10.1007/978-94-015-7799-1

Associated Press. 2025. "OpenAI's Sam Altman Warns of AI Voice Fraud Crisis in Banking." July 22; https://apnews.com/article/openai-ceo-sam-altman-fed-ad87262a4c1e71a0695ff6d06a2586f2

Bank for International Settlements and World Bank Group. 2016. Committee on Payments and Market Infrastructures and World Bank Group. *Payment Aspects of Financial Inclusion: Consultative Report*. Basel: Bank for International Settlements, September 2015. https://www.bis.org/cpmi/publ/d133.pdf

Bank for International Settlements. 2024. Carstens, Augustin. "Stronger Together: The Benefits of International Cooperation." September 3; https://www.bis.org/speeches/sp240903.htm

Bloomberg. 2025. Aldrick, Philip. "Bank of England's Bailey Says Deregulation Can Be Bad for Growth." February 11, 2025. https://www.bloomberg.com/news/articles/2025-02-11/bank-of-england-s-bailey-says-deregulation-can-be-bad-for-growth

Badawi, Adam. 2025. "The First Crypto President: Presidential Power and Crypto Currency Markets During Trump's Second Term (2025-2029)"; https://www.researchgate.net/publication/398313468_The_First_Crypto_President_Presidential_Power_and_Cryptocurrency_Markets_During_Trump's_Second_Term_2025-2029 https://doi.org/10.48550/arXiv.2512.03189

Bilenko, Michael Yuryevich. "Learnable Similarity Functions and Their Application to Record Linkage and Clustering." Mikhail Yuryevich Bilenko, UT Austin, 2006; available here: https://github.com/dedupeio/dedupe

Bouri, Elie, Elham Kamal, and Harald Kinateder. "FTX Collapse and Systemic Risk Spillovers from FTX Token to Major Cryptocurrencies." *Finance Research Letters* 56, no. C (2023); https://doi.org/10.1016/j.frl.2023.104099

Budish, Eric. 2025. "Trust at Scale: The Economic Limits of Cryptocurrencies and Blockchains." *Quarterly Journal of Economics*, Volume 140, Issue 1 (February), Pages 1–62; https://doi.org/10.1093/qje/qjae033

CBS News. 2025. "Trump Signs GENIUS Act Crypto Bill." July 18. https://www.cbsnews.com/news/trump-signs-GENIUS-act-crypto-bill/

CCN. 2025. "3 Key Bills US Investors Watch Crypto Week 2025." https://www.ccn.com/education/crypto/3-key-bills-us-investors-watch-crypto-week-2025/

Cliff, A. D., & Ord, J. K. 1981. *Spatial Processes: Models & Applications*. London: Pion.

Congressional Research Service. 2023. "Remittances: Background and Issues for the 118th Congress," May 10; https://www.congress.gov/crs-product/R43217

Congressional Research Service. 2025. *The State of Campaign Finance Policy: Recent Developments and Issues for Congress*. (CRS Report No. R41542, July 29.

CRN. 2025. "SEC Drops Remaining Claims Against SolarWinds Over 2020 Hack." https://www.crn.com/news/security/2025/sec-drops-remaining-claims-against-solarwinds-over-2020-hack

Desilver, Drew. 2022. "The Polarization in Today's Congress Has Roots That Go Back Decades." *Pew Research Center*, March 10, 2022. https://www.pewresearch.org/short-reads/2022/03/10/the-polarization-in-todays-congress-has-roots-that-go-back-decades/

Federal Deposit Insurance Corporation. 2023. *2023 FDIC National Survey of Unbanked and Underbanked Households*. Washington, DC: FDIC, 2023. https://www.fdic.gov/household-survey.

Federal Reserve Bank of Atlanta. 2025. "Geospatial Heterogeneity in Inflation: Market Concentration Story." Working Paper 2025-15, November 6; https://www.atlantafed.org/-/media/documents/research/publications/wp/2025/11/06/15-geospatial-heterogeneity-in-inflation-market-concentration-story.pdf

Ferguson, Thomas. 2025. "How Climate Denial Is Fueling a U.S. Homeowners Insurance Crisis—and Risking a 2008-Style Financial Meltdown." Institute for New Economic Thinking; https://www.ineteconomics.org/perspectives/blog/how-climate-denial-is-fueling-a-u-s-homeowners-insurance-crisis-and-risking-a-2008-style-financial-meltdown

Ferguson, T., Jorgensen, P., & Chen, J. 2013. Party Competition and Industrial Structure in the 2012 Elections: *International Journal of Political Economy, 42*(2), 3-41; https://www.tandfonline.com/doi/pdf/10.2753/IJP0891-1916420201

Ferguson, Thomas, Paul Jorgensen, and Jie Chen. 2020. "How Much Can the U.S. Congress Resist Political Money? A Quantitative Assessment." Institute for New Economic Thinking Working Paper No. 109; https://doi.org/10.36687/inetwp109

Ferguson, Thomas, Paul Jorgensen, and Jie Chen. 2022. "How Money Drives US Congressional Elections: Linear Models of Money and Outcomes." *Structural Change and Economic Development* Vol. 61, June, 527-45; https://doi.org/10.1016/j.strueco.2019.09.005

Ferguson, Thomas, Paul Jorgensen, and Jie Chen. 2022. "Bankman-Fried, Political Money, and the Crash of FTX." *Institute for New Economic Thinking*, November. https://www.ineteconomics.org/perspectives/blog/bankman-fried-political-money-and-the-crash-of-ftx.

Fellegi, I. P., & Sunter, A. B. (1969). A Theory for Record Linkage. *Journal of the American Statistical Association, 64(328), 1183–1210.* https://doi.org/10.1080/01621459.1969.10501049

Financial Times. 2025. "UBS Considers Moving to the U.S. to Escape Stricter Oversight." 2025. https://www.ft.com/content/fc270cb0-a7ad-4fc9-823f-c76c4e352034

Forbes. 2025. Alexander, Dan. "Presidency Boosts Trump's Net Worth by $3 Billion in a Year." *Forbes*, September 9; https://www.forbes.com/sites/danalexander/2025/09/09/presidency-boosts-trumps-net-worth-by-3-billion-in-a-year/

Holland & Knight. 2024. "Court in SolarWinds Case Blows Down SEC's Cyber Enforcement Authority." July; https://www.hklaw.com/en/insights/publications/2024/07/court-in-solarwinds-case-blows-down-secs-cyber-enforcement-authority

Institute for New Economic Thinking. 2025. "Red Tech and American Politics: Nick French Interviews Thomas Ferguson." https://www.ineteconomics.org/perspectives/blog/red-tech-and-american-politics-nick-french-interviews-thomas-ferguson

International Consortium of Investigative Journalists. 2025. "The Coin Laundry." https://www.icij.org/investigations/coin-laundry/about-coin-laundry-investigation-cryptocurrency/

International Monetary Fund. 2023. "Pix: Brazil's Successful Instant Payment System." *IMF eLibrary*, 2023. https://www.elibrary.imf.org/view/journals/002/2023/289/article-A004-en.xml

Jorgensen, Paul. 2013. "Pharmaceuticals, Political Money, and Public Policy: A Theoretical and Empirical Agenda." *Journal of Law, Medicine, and Ethics* 41(3): 561-570. https://doi.org/10.1111/jlme.12065

Lalisse, Matthias. 2024. "The Long Goodbye: Mitch McConnell and Big Money Politics." Institute for New Economic Thinking; May 16; https://www.ineteconomics.org/perspectives/blog/the-long-goodbye-mitch-mcconnell-and-big-money-politics

Kaur, Pam. 2025. "Why The OCC's Crypto Bank Charter Push Looks Deregulatory." *Forbes*, December 12; https://www.forbes.com/sites/pamkaur/2025/12/12/why-the-occs-crypto-bank-charter-push-looks-deregulatory/

Ma, Yiming, Zheng, Yao, and Zhang, Anthony. 2025. "Stablecoin Runs and the Centralization of Arbitrage." University of Chicago: Becker-Friedman Institute. Working Paper No. 2025-76; https://bfi.uchicago.edu/wp-content/uploads/2025/06/BFI_WP_2025-76.pdf http://doi.org/10.3386/w33882

Mian, Atif, Amir Sufi, and Francesco Trebbi. 2010. "The Political Economy of the U.S. Mortgage Default Crisis." *American Economic Review* 100, no. 5 (2010): 1967–98. https://doi.org/10.1257/aer.100.5.1967

Mian, Atif, Amir Sufi, and Francesco Trebbi. 2013. "The Political Economy of the Subprime Mortgage Credit Expansion." *Quarterly Journal of Political Science* 8, no. 4 (2013): 373–408. https://doi.org/10.1561/100.00012036

Morris, G. Elliott. 2023. "Polarization in the US House: The Data." *Substack*, January. https://www.gelliottmorris.com/p/polarization-in-the-us-house-the.

New York Times. 2025a. Copeland, Rob. "Behind Wall Street's Abrupt Flip on Crypto." August 13, 2025. https://www.nytimes.com/2025/08/13/business/wall-street-banks-crypto-stablecoins.html

New York Times. 2025b. Protess, Ben, Fuller, Andrea, LaFroniere, Sharon, Hughes, Seamus. "The S.E.C. Was Tough on Crypto. It Pulled Back After Trump Returned to Office." December 14; https://www.nytimes.com/2025/12/14/us/politics/sec-crypto-firms-trump-investigation.html

Nokken, Timothy P., and Keith T. Poole. 2021. "Congressional Voting Data." *Voteview*. https://voteview.com/articles/data_help_members.

Parramore, Lynn. 2025. "Ex-CISA Official Warns: We've Gutted Cybersecurity—a Gift to Iran, China, and Russia." Institute for New Economic Thinking; https://www.ineteconomics.org/perspectives/blog/ex-cisa-official-warns-weve-gutted-cybersecurity-a-gift-to-iran-china-and-russia

Reuters. 2025. Howcroft, Elizabeth. "Tether's Stablecoin Downgraded in Weak S&P Assessment." November 26, 2025. https://www.reuters.com/business/finance/tethers-stablecoin-downgraded-weak-sp-assessment-2025-11-26/.

Sakellariadis, John. 2025. "Acting CISA Director Failed a Polygraph Test. Career Staff Are Now Under Suspicion." Politico, Dec. 21; https://www.msn.com/en-us/news/politics/acting-cisa-director-failed-a-polygraph-career-staff-are-now-under-investigation/ar-AA1SMbly?ocid=BingNewsSerp

Schwarcz, Daniel, and Josephine Wolff. 2025. "Why Liability and Insurance Won't Save AI: Lessons From Cyber Insurance." *Lawfare*, September 9; https://www.lawfaremedia.org/article/why-liability-and-insurance-won-t-save-ai--lessons-from-cyber-insurance

Scott, Brett. 2022. *Cloudmoney*. New York: Harper Business.

Silverman, Jacob. 2025. *Gilded Rage: Elon Musk and the Radicalization of Silicon Valley*. London: Bloomsbury Continuum. https://doi.org/10.5040/9781399420020

The Quantum Insider. 2024. Swayne, Matt. "Chinese Quantum Computing Advance Shows Progress, Innovation, but Not an Imminent Threat to Encryption." October 22, 2024. https://thequantuminsider.com/2024/10/17/chinese-quantum-computing-advance-shows-progress-innovation-but-not-an-imminent-threat-to-encryption/

U.S. Securities and Exchange Commission. 2023. *Complaint: SEC v. SolarWinds*. https://www.sec.gov/files/litigation/complaints/2023/comp-pr2023-227.pdf.

Wall Street Journal. Rosenbush, Steve. 2024. "Q-Day Is Coming: It's Time to Worry About Quantum Security," October 9; https://www.wsj.com/articles/q-day-is-coming-its-time-to-worry-about-quantum-security-ca88c576

Washington Post. 2020. Harwell, Drew and Macmillan, Douglas. "SolarWinds Russia Breach Stock Trades." December 15; https://www.washingtonpost.com/technology/2020/12/15/solarwinds-russia-breach-stock-trades/

Washington Post. 2025. Joseph Menn. Shutdown Guts U.S. Cybersecurity Agency at Perilous Time. October 2; https://www.washingtonpost.com/technology/2025/10/02/cisa-shutdown-cybersecurity/

White, Geoff. 2024. *Rinsed: From Cartels to Crypto: How the Tech Industry Washes Money for the World's Deadliest Crooks*. London: Penguin Books.

Wilmarth, Arthur E., Jr . 2025a. "Time Bomb: How Uninsured Stablecoins and Crypto Derivatives Threaten Financial and Economic Stability." Institute for New Economic Thinking blogpost; https://www.ineteconomics.org/perspectives/blog/time-bomb-how-uninsured-stablecoins-and-crypto-derivatives-threaten-financial-and-economic-stability

Wilmarth, Arthur E., Jr. 2025b. “The Looming Threat of Uninsured Nonbank Stablecoins,” 50 Delaware Journal of Corporate Law 3-135; https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5272859

Wired. Zetter, Kim. 2023. "The Untold Story of SolarWinds: The Boldest Supply-Chain Hack Ever." May 2; https://www.wired.com/story/the-untold-story-of-solarwinds-the-boldest-supply-chain-hack-ever/

Wired. Willems, Adam. 2025a. “The Loophole Turning Stablecoins into a Trillion-Dollar Fight.” September 3; https://www.wired.com/story/GENIUS-act-loophole-stablecoins-banks/

Wired. Newman, Lily. 2025b. “Fears Mount That US Federal Cybersecurity Is Stagnating—or Worse.” December 31; https://www.wired.com/story/expired-tired-wired-federal-cybersecurity/

Notes

We are grateful for valuable advice and assistance to Philip Basil, Peter Bofinger, Amanda Fischer, Pia Malaney, Perry Mehrling, Lynn Parramore, and Arthur Wilmarth. An earlier version of this paper was presented at a Better Markets symposium in September, 2025.

[1] The literature is large, if uneven. Among the most illuminating is (Bouri, et al., 2023).
[2] Here, again, the literature is very large, though also repetitious, including substantial discussions in the *Financial Times*, the *New York Times*, CBS News, and other sources. A useful review is Silverman (2025), esp. Chapters 13 and 18. Badawi (2025) is also very helpful.
[3] The 2024 legislative measures were not identical to the 2025 bills; the closest was the May vote on the Financial Innovation and Technology for the 21st Century Act. See the discussion below. A later vote in July to uphold President Biden's veto of a Congressional effort to overturn a ruling by the Securities and Exchange Commission on crypto accounting received much press attention, but does not brook comparison, since it involved a highly partisan repudiation of a Democratic president.
[4] Senate Democrats are split on crypto. We are tracking contributions there, too, but as discussed below, the small number of cases complicate analyses. It does not help that states are typically much larger than congressional districts, so that, for example, findings like ours on the percentage Hispanics or Latinos in districts might not show at all.
[5] Measuring "ideology" in Congress is a contentious cottage industry, but we lack the space for a detailed discussion. See the references and discussion in our Statistical Appendix.
[6] See the discussion of (Mian, Sufi, and Trebbi, 2010) and (Mian, Sufi, and Trebbi, 2013) in (Ferguson, Jorgensen, and Chen, 2020), esp, 55ff. More generally, (Ferguson, Jorgensen, and Chen, 2022) and (Ferguson, Jorgensen and Chen, 2013); (Lalisse, 2024).
[7] In the jargon of statistics, each legislator becomes their own control. It is worth cautioning that the statistical nature of the argument means that assertions about particular individuals need always to be qualified; this is different than for the group as a whole. The point can be instructively compared with the use of individual test scores to advise specific students: the chance of error in individual cases is real. Note that a longer version of our Congress paper appeared as an essay in Rochon, Louis-Philippe, and Borgrine, Hassan. 2020. *Credit, Money and Crises in Post-Keynesian Economics.* Cheltenham: Edward Elgar, 152-205.
[8] The present study does not share all these advantages. We want, accordingly, to acknowledge the possibility that some other unexamined variables may also be operating. As outlined in our Statistical Appendix, we have taken pains to check for multicollinearity and scrutinized a lot of other variables, such as median household district income, before arriving at our model.
[9] See, inter alia, (Scott 2022); (White, 2024); (Ma, Zheng, and Zhang, 2025) is eye opening on arbitrage issues and literature.
[10] Which begs the question of what the federal policy response will be in crisis. In all likelihood, the stablecoin industry will seek some form of federal lender of last resort protection. On this, more some other time. In the meantime, see the recent decision of the Office of the Comptroller of the Currency to grant some crypto firms conditional national trust bank charters.
[11] Which is obviously at the root of many of the conflicts between traditional finance and crypto.

[12] Even casual web searches will reveal many references to remittances in crypto debates by members of both parties, such as Tom Emmer or Ro Khanna. See, e.g., (Congressional Research Service 2023)
[13] Reports from BIS and World Bank working groups have repeatedly emphasized how accounts for basic services can be designed to have very low know your customer costs. See., e.g., Committee on Payments and Market Infrastructures and World Bank Group (2015). Payment aspects of financial inclusion. Consultative report, September 2015. Bank for International Settlements. Available at: https://www.bis.org/cpmi/publ/d133.pdf The problem is that the profitability of such accounts is relatively low. See below.
[14] The numbers change drastically when one distinguishes between "under-banked" people with very rudimentary access and no access at all. See the report.
[15] Cf. BIS and World Bank Group Committee on Payments and Market Infrastructures, Payment Aspects of Financial inclusion, Consultative report, September 2015. Bank for International Settlements. Available at: https://www.bis.org/cpmi/publ/d133.pdf Note that the report is copyrighted 2016.
[16] Cf. Wilmarth: "Using stablecoins held in private digital wallets makes it significantly harder for private parties and government authorities to track the identities of crypto traders, compared to payments made with fiat currency that are routed through traditional financial institutions."
[17] Excellent reviews of many cases of breathtaking scale and audacity are also presented in the International Consortium of Investigative Journalism, "The Coin Laundry."
[18] One is said to be coming – eventually.
[19] No published analysis of Solar Winds does justice to the range of concerns the incident raises. But see the original SEC complaint, noting the subsequent history of the case; and the Washington Post discussion of stock sales. Most concerning of all, though, should perhaps be instances in which whistles did not blow despite signs of anomalous behavior. See, e.g., the discussion in (Wired 2025b).
[20] We tested for state fixed effects for both outcomes; we did not find these.
[21] See https://voteview.com/articles/data_help_members
[22] Just over a fifth (21.6%) of all Democrats received no contributions; 39% took in less than five thousand dollars; with just over a half (55.4%) garnering less than 10 thousand. Twenty percent received $100,000 or more.
[23] For example, in the 2022 cycle FTX and its trading affiliate Alameda Research, along with a circle of straw donors accounted for roughly a half or more of the industry's total spending, using in some cases pop ups. The 2024 election displayed an analogous pattern. A compact network of super PACs funded by crypto intermediaries Coinbase and Ripple and a few others, notably A16z. They, too, sometimes used intermediaries and all are in our tabulations.
[24] We provide the full list of PACs included in the analysis here: Coinbase Innovation PAC (C00804179), Web3 Forward (C00804187), GMI PAC (C00788679), HODLPAC (C00733691), Chamber of Digital Commerce PAC (C00567412), Blockchain Association PAC (C00824896), Payward Ventures/Kraken PAC (C00845651), Galaxy Digital LLC PAC (C00879452), Keep Startups in America (C00852657), Circle Internet Group PAC (C00894055), American Dream Federal Action (C00809020), Crypto Innovation (C00804732), Protect Our Future PAC (C00801514), Fair Shot USA (C00821041), Protect Progress (C00848440), Fairshake (C00835959), Defend American Jobs (C00836221), Crypto Freedom PAC (C00816892), DAO For America (C00811711), CryptoPAC, Inc. (C00787085), Financial Freedom PAC (C00810002), Crypto Foundation of America (C00806729), Stand with Crypto Alliance, Inc.

PAC (C00876631), Crypto/Bitcoin USA Super PAC (C00882472), Crypto for US (C00884429), Crypto and Data Center PAC (C00707356), US Federal Blockchain PAC (C00845628), Blockchain Freedom PAC (C00851873), American Blockchain PAC (C00779819), Digital Asset PAC (C00789115), Digital Asset Alliance (C00793604), Blockchain of Command (C00739920), American Blockchain & Cryptocurrency PAC (C00736934), Bitcoin Voter PAC (C00880518), Make Bitcoin Great Again (C00881250), Bitcoin Freedom PAC (C00822775). Two additional PACs—Bitcoin Freedom PAC (C00822775, Cheri Beasley) and Commonwealth Unity Fund, (C00875856, John Deaton) did double duty as single-candidate Super PACs, with Commonwealth Unity Fund spending $2.1 million against Elizabeth Warren in the 2024 Massachusetts Senate race. In both cases, each and every one of their its donors were identifiable as large crypto investors (38 percent of CUF's operating money coming from a $1 million contribution from Ripple Labs Inc.), so we classified these as crypto PACs directly.

[25] See, e.g., (Jorgensen 2013) for a discussion.

[26] LLMs are rapid to initialize on tasks where substitution with a classical computer program would require reams of code and processing routines to approximate the intuition of a human annotator performing the same task. LLMs (the largest ones in any case) are more than capable of these modest leaps of intelligence, and can replace the annotator and the classical program alike in settings where throughput requirements are not too extreme. Unfortunately, scaling requirements in modern-day FEC datasets often fall in the latter category, so while the vaunted virtues of the Large Language Model are quite real, both speed and accuracy considerations compel the use of stacks of other methods in the main.

[27] Which means that for any give analysis, one wants to attend both to the dates of the campaign finance reports that were used and also when analysis stopped, since that improves over time.

[28] Our deduplication pipeline runs end-to-end on a full cycle of 2024 campaign finance data in about 24 hours, using bulk datasets consisting of some 75 million transactions with about 8 million unique donor identifiers. The number of pairwise comparisons (without the shortcut of data-blocking) is on the order of tens of trillions.

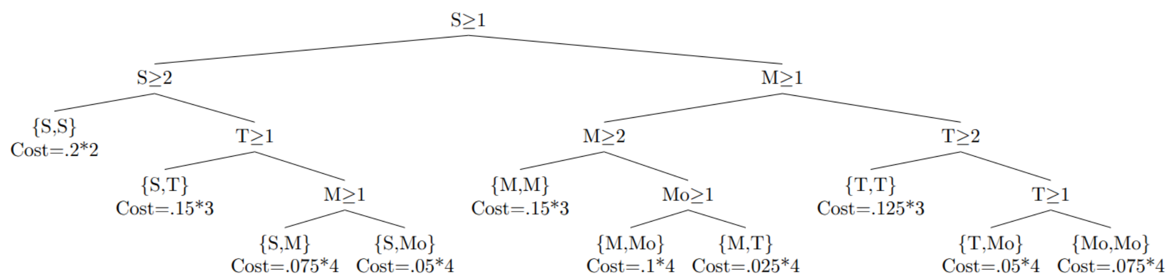
29 For instance, there are 7 categorical entity types in FEC data, INDividual, CANdidate, COMmmittee, ORGanization, PAC, and PTY (party). We add an eighth category (UORG) for noisy entity types generated using the probablepeople parser for entries from IRS 527 filings, in which entity type is not indicated (we must guess it, and errors do occur. Strict equality coding (a single feature for type1 == type2) precludes the model from conditioning the handling of other features on the specific entity types involved in the combination, e.g. to require more stringent evidence from location for individuals than organizations, the latter being more mobile. The default solution in dedupe is an indicator coding of the possible conjunctive variable values, with 8C2-1=27 new variables (with one pair-category offloaded to the logistic model's constant) to capture each unordered conjunction (Individual-to-Individual, Individual-to-Organization, etc.). When greater expressivity is available in the form of nonlinear decision boundaries (RF), the feature combinations can instead be expressed with one sum-indicator per category—e.g. (type1 == cat_i) + (type2 == cat_i) for each of the 8 categories—with features iteratively consulted when building trees, the resulting trees being shallower as a consequence. To see that the trees are shallower, consider the following construction. Suppose our algorithm needs to determine which categorical bin it is in in order to maximize informativeness of the numerical features used in subsequent steps. As an instructive example, take a simpler (four-class) problem involving vehicle collisions, e.g. to predict the probable survival rates of the participants in a new collision. Other features may be involved, but suppose first that the algorithm must determine, by

examining features, the type of two vehicles, while also minimizing the number of computational steps taken. The vehicle-type conjunctions are distributed as the upper-triangle of the following normalized contingency table:

	mini	**truck**	**sedan**	**moto**	**total**
mini	.15	.025	.075	.1	.35
truck	.025	.125	.15	.05	.35
sedan	.075	.15	.2	.05	.475
moto	.1	.05	.05	.075	.275

The idea behind the probabilities is that sedans are most common, and minis (a small car, e.g. mini cooper or smart car) and motorcycles (moto) are likely to give trucks wide berth to minimize personal hazard (similarly for sedans). We consider two encoding strategies: one feature per (unordered) conjunction ("conjunctive", 10 features), and one feature per unary category ("disjunctive", 4 features). In the disjunctive scheme, the value of the feature is the sum of category membership for each vehicle (2 if both are of the same class, 1 if one of the two is of the class, 0 if neither). Conceptually, the disjunctive scheme allows steps to be ordered in such a way that the feature-checking order is conditioned on the results of previous sampling steps—which are informative to the degree that crash incidence for two vehicles is non-independent—allowing that order to be optimized row-wise, and also systematically exclude cells inconsistent with the results of prior feature tests. For note that, in the conjunctive coding scheme, one can do no better than to sort the conjunctive features inversely to the probability of each conjunction, with expected cost E[cost] = min(p)*9 + $\sum_{i=[1,9]} p_i * i$ (the last class is identified by elimination at no additional cost). Any modification at all to the indicated ordering is defective, since it increases the mass on a larger cost factor i, while reducing that on a lower factor j (in general, the cost is minimized by sorting so as to minimize the correlation between the cost vector [1,…,9,9] with the probability vector, which is strictly the inverse-probability sort). The expected cost under our scenario (10 conjunctive features) is exactly: .2*1 + .15*2 + .125*3 + .1*(4+5+6) + .075*(7+8) + .05*9 + .025*9 = 4.175 steps on average, with a minimum of one step and a maximum of nine steps required.

The compact RF with four features (disjunctive) would instead iterate by first selecting the class with highest single-vehicle mass (sedan), then iterating in sequence on the most likely vehicle type given that a sedan was involved in the collision (disposing of high conditional-mass branches earlier), while also paring branches inconsistent with the outcome of previous sampling steps. An efficient procedure yields the following tree. Left branches indicate a result of 1 on the feature test, with leaf nodes annotated for cost (tree depth=number of steps required for identification):



Connectedness of feature test results with cell exclusions is the key property exploited. Involvement of a sedan (with row-wise mass .475 collisions) being the highest-probability category, the algorithm would iterate on features in order of highest probability along that branch, disposing of high-mass nodes first (at lower cost). The minimum is two (identification off the diagonal requiring at least two feature checks) and a maximum of four steps if a sedan was involved in the collision:

Branch cost (*sedan* ≥ 1): .2*2 + .15*3 + .075*4 + .05*4 = 1.35

The right branch from this root node of the decision tree terminates only in leaves not involving S (as this feature need not be queried again). We repeat the above process for the row with the highest residual mass having excluded *sedan* (*mini*, 52% of remaining samples), with a minimum leaf cost of three steps (having expended one step excluding S) and a maximum of four. The rightmost subtree (root T≥2) results from a local optimization with cost savings of .05 from disposing of the higher-mass *truck-truck* branch earlier than lower-mass *moto-moto*. The total expected cost is E[cost] = .2*2 + (.15 + .15 + .125)*3 + (.075 + .05 + .1 + .025 + .05 + .075)*4 = 3.175, with a worst-case cost of 4 and average-case performance that is 20.6% better than the conjunctive scheme (shallower trees in principle, and also on average). Further savings are possible with a frequency-aware binary encoding of the combinations, but at cost then of post-hoc interpretability and robustness to drift.

We have engineered each feature to maximize model expressiveness while also minimizing computational cost for the pair-scoring step. Given that this component of the record linkage pipeline is generally the most time-consuming, optimizations of this sort can make great differences for program runtime.