Is Fraud Contagious? Social Connections and the Looting of COVID Relief Programs*

John M. Griffin[†] Samuel Kruger[‡] Prateek Mahajan[§]

August 3, 2024

Abstract

Fraud indicators in the Paycheck Protection Program (PPP) COVID relief program are highly geographically concentrated. Areas with high PPP fraud also have heightened indicators of suspicious Economic Injury Disaster Loan (EIDL) advances and unemployment insurance claims. Zip codes and counties with high rates of suspicious PPP loans exhibit strong social connections to one another with evidence of fraud spreading over time through social connections. Additionally, individuals in suspicious social media groups have higher rates of PPP fraud, and socially connected zip codes frequently use the same specific FinTech lenders and EIDL agents, consistent with social connections influencing detailed loan decisions.

JEL classification: G21, G23, G28, H12
keywords: Fraud, Social Media, Government Spending

*We thank Jim Conklin, Mike Faulkender, William Fuchs, Caitlin Gorback, Christoph Herpfer, Elena Loutskina, Lin Peng, Alex Priest, Marius Ring, Jacob Sagi, Laura Starks, Constantine Yannelis, Eric Zwick, seminar participants at the University of North Carolina, the University of Utah, and the Federal Reserve Bank of Dallas, anonymous industry participants, and others for helpful comments. Some of the material in this paper was initially included in a related paper and benefitted from comments from seminar and conference participants at the American Real Estate and Urban Economics Association (AREUEA) national conference, Federal Reserve Bank of Dallas, the Financial Intermediation Research Society (FIRS) conference, the Office of the Comptroller of the Currency (OCC) Emerging Risks Symposium, the University of Kentucky Finance Conference, the University of Pennsylvania, the University of Texas at Austin, and the University of Utah. We also thank the University of Texas at Austin and the Integra Research Group for research support. Griffin is an owner of Integra Research Group and Integra FEC, which engage in research, financial consulting, and recovery on a variety of issues related to the investigation of financial fraud.

†McCombs School of Business, University of Texas at Austin. John.Griffin@utexas.edu
‡McCombs School of Business, University of Texas at Austin. Sam.Kruger@mccombs.utexas.edu
§McCombs School of Business, University of Texas at Austin. Prateek.Mahajan@mccombs.utexas.edu

Is Fraud Contagious? Social Connections and the Looting of COVID Relief Programs

August 3, 2024

Abstract

Fraud indicators in the Paycheck Protection Program (PPP) COVID relief program are highly geographically concentrated. Areas with high PPP fraud also have heightened indicators of suspicious Economic Injury Disaster Loan (EIDL) advances and unemployment insurance claims. Zip codes and counties with high rates of suspicious PPP loans exhibit strong social connections to one another with evidence of fraud spreading over time through social connections. Additionally, individuals in suspicious social media groups have higher rates of PPP fraud, and socially connected zip codes frequently use the same specific FinTech lenders and EIDL agents, consistent with social connections influencing detailed loan decisions.

JEL classification: G21, G23, G28, H12
keywords: Fraud, Social Media, Government Spending

How does fraud spread, and to what extent is it contagious? In the $4.42 trillion spent on COVID relief programs, there is growing evidence of "an unprecedented amount of fraud," described at the ground level as "the biggest fraud in a generation."[1] What is behind this fraud? Did sophisticated cybercriminals exploit system weaknesses, or was the effort more grassroots? Were some individuals simply more likely to fudge numbers, or were they recruited? Did the fraud spread locally or over longer distances through social media and social connections? This paper investigates connections between fraud across programs and examines how fraud spread over time and across geographies, particularly through social connections. Overall, our results indicate that the looting of government programs can spread rapidly through social connections in today's digital world, and law enforcement and government administrators may need to be much more proactive in response.

We first identify stark geographic variation in Paycheck Protection Program (PPP) relief fraud. Using loan-level flags for suspicious loans developed by Griffin, Kruger, and Mahajan (2023), we find that while some counties have almost no PPP fraud, other counties have in excess of 30% of loans flagged as suspicious. Variation is even higher at the zip code level, with flag rates ranging as much as 5% to 45% within counties.

Is PPP fraud related to suspicious relief spending in other pandemic programs? Government audits have identified substantial fraud in pandemic unemployment insurance (UI) programs and in the Economic Injury Disaster Loan (EIDL) program, but these programs differ from the PPP in their target populations, application processes, and use of intermediaries. To the best of our knowledge, there has been no systematic academic evaluation of suspicious activity in the EIDL program. Recent work by Khetan et al. (2024) investigates fraud in pandemic unemployment insurance. We create measures for suspicious EIDL advances and excess UI claims. In particular, the EIDL Advance program created a potential incentive to inflate the number of employees being reported since applicants were eligible for an EIDL Advance grant of $1,000 per employee, up to a maximum of $10,000. We identify inflated employee reporting based on discrepancies between matched EIDL and PPP applications. Zip codes that have higher rates of advances with inflated jobs also have higher rates of suspicious PPP loans. Moreover, these zip codes also have significantly

[1] Mike Galdo, the U.S. Justice Department's acting director for COVID-19 Fraud Enforcement, quoted in Associated Press (2023) and former U.S. Attorney Matthew Schneider, quoted by NBC News (2022). See here for details on pandemic relief spending including COVID-19 unemployment benefits of $872 billion, the Paycheck Protection Program (PPP) totaling $793 billion, and the Economic Injury Disaster Loan (EIDL) program totaling $384 billion. The percentage of funds lost to fraud ranges from 12% to 33% in various government studies and audits, which are more fully described in Section 3.

higher rates of recipients receiving EIDL Advance grants, which did not need to be repaid, without receiving the corresponding loan component of the EIDL.[2] Counties with high rates of PPP fraud also show evidence of elevated unemployment insurance claims after UI was expanded in 2020, but not before; specifically, elevated claims in high PPP fraud counties begin precisely in March and April 2020 and persist until November 2020, despite no indication of differential economic shocks in these areas during the relevant time period.

Can the geographic concentration in pandemic fraud across programs be explained through social networks? To test this hypothesis, we use social connectedness between geographic areas, controlling for total number of users in each location, based on data constructed by Bailey et al. (2018a, 2020) from the Facebook social graph. Zip codes with high PPP fraud rates have disproportionately high Facebook friendships with other zip codes with high fraud, even when located in distant parts of the country. To distinguish social connections from physical proximity, we examine how suspicious PPP (and separately EIDL) lending in a zip code relates to suspicious lending rates in other zip codes that are socially and physically proximate to it. Social proximity to fraud is highly predictive of zip code fraud rates, and physical proximity is unrelated to the fraud rate in the zip code after controlling for the fraud rate of socially connected zip codes.

Does social proximity reflect homophily or other unobserved similarities between zip codes with high fraud rates? In particular, socially connected zip codes likely share similar characteristics such as race and income. The detailed zip code demographic variables in our baseline specifications control for homophily on important observed dimensions (income, poverty rates, population density, minority population, educational attainment, and pre-pandemic unemployment) and have little impact on the influence of social proximity. Based on the logic of Oster (2019), this suggests that omitted variable bias is likely to be small in our setting. We find similar effects of social proximity in subsamples of zip codes with below and above median levels of a wide range of demographic and economic characteristics, pointing to a broad-based effect of social proximity that spreads through social groups in general as opposed to an effect that is mainly driven by particular demographic or economic groups. We instrument for social proximity to suspicious lending using only distant zip

[2]The EIDL program had two components: 1) an advance grant, which provided businesses $1,000 per employee, up to 10 employees, and did not need to be repaid, and 2) a more traditional loan program that provided loans for up to $2 million, which had to be repaid. Zip codes with the highest rates of PPP fraud have 60% more EIDL Advance grant recipients who did not take out a corresponding loan.

codes (e.g., over 500 miles away), which has the benefit of avoiding any omitted variables related to local areas such as in-person exposure to fraud. The results are extremely similar. To directly test for homophily, we also create an analogous measure of the demographic similarity between zip codes and find that the influence on social proximity is unaffected by the inclusion of this measure.

Do culture, historic rates of identity theft, historic levels of suspicious financial activity, and social capital play a role in spreading suspicious lending? We examine variables found in previous papers to be associated with the geography of fraud. These measures provide little incremental explanatory power in explaining the geography of PPP fraud and including them does not affect the strong relation between social proximity and suspicious lending rates. Using measures of social capital from Chetty et al. (2022a,b), we find that clustering and civic connections are positively related to zip code fraud rates, while volunteering is negatively related to fraud. These variables have little effect on social proximity, and indicate that social interactions may have had a role in spreading fraud.

Can social networks predict how fraud spread over time? We first consider matched zip codes with identical fraud rates in round 1 of the PPP program but with different social connections. The zip codes with strong social connections to others with PPP fraud have much faster growth in fraud. In a regression context, zip code-level fraud rates in 2021 are strongly predicted by the 2020 fraud rates of its socially proximate zip codes even after controlling for the zip code's past fraud, and the effect again subsumes physical proximity.

What is the mechanism through which social connections transmit fraud? Do members of social media groups promoting questionable activity receive more fraudulent PPP loans? We search for and identify 136 public social media groups discussing pandemic relief programs on a large social media platform. We use unique names to match over 17,600 members of these groups to loan-level PPP data. Individuals involved with groups using words indicative of pushing fraud opportunities (such as "document," "method," and "sauce") exhibit substantially higher rates of suspicious PPP loans compared to members of more benign groups.

To further understand the mechanism through which social connections influence pandemic relief fraud, we examine whether social connections explain specific decisions such as specific loan features and which lender and agent a borrower uses. Clusters of loans with identical features such as the same lender, the same county, the same industry, the same loan amount, and the same number of employees are potentially indicative of fraudulent applicants submitting the same loan information (Griffin, Kruger, and Mahajan, 2023), consistent with applicants sharing how-to information for obtaining fraudulent PPP loans (frequently referred to as secret "sauce" and "methods" in social media posts). In addition to being concentrated in particular counties, these clusters of identical loans are also more common in socially connected counties. Members of social media groups also tend to use the FinTech PPP lender that was most discussed in their group. More generally, PPP recipients frequently use the FinTech platform which has the highest usage in socially connected zip codes. In contrast, social connections do not explain lender usage for the two largest traditional banks, both of which have low suspicious loan rates. These relations highlight that social connections influenced specific decisions such as identical loan size, reported employees, reported industry, and what lender to use; which is not what one would expect if social connections merely capture homophily or other omitted similarities across zip codes.

Switching to the EIDL program, in response to our Freedom of Information Act (FOIA) request, the SBA released data on over 450,000 loans where an agent facilitated the EIDL loan. When an agent operates in multiple zip codes, we find that the social connections between zip code pairs predict which zip codes the agent is active in. Interestingly, fraud rates vary significantly more across agents than expected based on simulations. In particular, more agents have zero EIDL fraud than the simulations would predict, but other agents appear to specialize in facilitating suspicious loans and have significantly higher fraud rates than we observe in the simulations. Fraud rates on an agent's previous loans are also highly predictive of subsequent fraud. This demonstrates another mechanism through which social networks seemingly facilitated the spread of pandemic fraud.

To understand whether our findings have potential longer-term implications, we examine whether social media users in groups related to suspicious pandemic relief continue to be active in groups related to suspicious financial activity in 2022–23. Social media users who were in pandemic relief groups with high PPP fraud rates are significantly more likely to also be in groups that discuss suspicious financial activity in 2022–23. Additionally, social media users in pandemic relief groups with high PPP fraud rates are also more likely to be in groups discussing the Employee Retention

Tax Credit (ERC), which has been flagged by the IRS and prominent media outlets for its susceptibility to fraudulent claims (IRS, 2023; Wall Street Journal, 2023a). Using data from Google Trends, we also find that areas of the country which had high PPP fraud rates in 2020–21 have much higher levels of search activity related to the ERC in 2022–23.

Our paper contributes to four main literatures. First, it illustrates how fraud spreads.[3] Glaeser, Sacerdote, and Scheinkman (1996) model how social interactions can explain the large geographic variation in crime rates across cities and neighborhoods, with social interactions being more important for nonviolent crimes. Bikhchandani, Hirshleifer, and Welch (1998) model how even a small shock can lead to a cascade in social norms, including norms related to crime, due to individuals learning from the behavior of others.[4] Dimmock, Gerken, and Graham (2018) show that connections among individuals working in the same office spread financial misconduct. Levi (2008) argues that organized crime is becoming more decentralized over time. Chetty, Friedman, and Saez (2013) find that geographic variation in income manipulation incentivized by the Earned Income Tax Credit program slowly grew and spread geographically over time from one percent in 1996 to around three percent in 2009. Healthcare fraud by home health agencies also appears to spread slowly over time through patient-sharing networks (O'Malley, Bubolz, and Skinner, 2023). Our findings indicate that the spread of fraud through social networks may grow much more rapidly than the geographic spread documented by the prior literature.

Second, our paper relates to a broad literature on the influence of social media, networks, and interactions. Social media may increase polarization (Levy, 2021) and can quickly spread false information (Vosoughi, Roy, and Aral, 2018) including conspiracies, health misinformation (Del Vicario et al., 2016; Suarez-Lledo and Alvarez-Galvez, 2021), and vaccine hesitancy (Puri et al., 2020). Because of the homophily of social networks and biases in information diffusion, social media can act like an echo chamber (Cinelli et al., 2021), which can foster hate crimes (Müller and Schwarz, 2020) and even recruitment into gangs (Décary-Hêtu and Morselli, 2011). Social networks and interactions have also been shown to influence a variety of financial activities (Hirshleifer, 2020) including

[3] This also relates to a broader literature on detecting (Silverman and Skinner, 2004; Van Vlasselaer et al., 2017; Shekhar, Leder-Luis, and Akoglu, 2023) and deterring fraud (Eliason et al., 2021; Shi, 2023) fraud.

[4] Baker and Faulkner (2003) detail how Ponzi schemes spread through social networks based on existing investors influencing others in their social circles to invest. Ponzi schemes spread through affinity groups (Deason et al., 2015) and related social groups (Nash, Bouchard, and Malm, 2013), and result in a lack of trust (Gurun, Stoffman, and Yonker, 2017). Similarly, Holzman, Miller, and Williams (2021) find that 296 SEC announcements of accounting misconduct are followed by increases in local financial crime, potentially due to eroding social norms.

retirement plan participation (Duflo and Saez, 2003), claiming of income tax credits (Wilson, 2022), stock market participation and trading (Hong, Kubik, and Stein, 2004; Fang and Seasholes, 2004), home purchases and beliefs (Bailey et al., 2018b,c), mortgage refinancing (Maturana and Nickerson, 2018), strategic mortgage default (Guiso, Sapienza, and Zingales, 2013), the use of online lending marketplaces (Allen, Peng, and Shan, 2024), and the use of specific banks (Cramer and Koont, 2021).[5] We extend this literature by examining the role that social networks play in transmitting and amplifying fraudulent financial activity.

Third, there is a literature examining the role of peer effects in crime. Sutherland's (1939) differential association theory argues that the underling conditions for crime are rooted in conflicts over norms, values, and interests (Matsueda, 1988). In their survey of the criminology literature, McGloin and Thomas (2019) state that they believe there is ample causal evidence on the question of whether peers influence delinquent behavior (e.g., McGloin, Thomas, and Sullivan 2019), and encourage more work "digging into questions about mechanisms and process" including "more complex questions such as how, when, and among whom do peers influence behavior." They note that most of the literature studying social interactions and deviant behavior is based on close friend groups and/or influence among adolescences and that we should seek to understand peer influence mechanisms in other situational and virtual settings.

Fourth, we contribute to further understanding the impact of COVID relief spending. Chetty et al. (2023) find that the cost of each job saved by the PPP was $377,000, and Autor et al. (2022) find costs of $170,000 to $257,000 per job. Granja et al. (2022) find small effects of the PPP on employment.[6] Griffin, Kruger, and Mahajan (2023) develop and cross-validate loan-level measures of suspicious PPP lending and find that fraudulent lending concentrates in FinTech lenders. We add to this literature by creating indicators for suspicious EIDL advances and excess unemployment insurance claims, and by examining how weaknesses in all these COVID relief programs may have

[5] Kuchler and Stroebel (2021) provide an overview of the social finance literature. Other networks such as those between (Acemoglu et al., 2012) and within Giroud and Mueller (2019) firms have also been shown to be important channels in spreading and amplifying shocks to economic activity.

[6] In contrast, Faulkender, Jackman, and Miran (2021) find that the program was much more effective with an average cost per job saved of $28,000. Erel and Liebersohn (2022), Chernenko and Scharfstein (2024), and Howell et al. (2024) find that minority-owned businesses were less likely to receive PPP loans from traditional lenders and that FinTech lenders helped close the funding gap. Regarding the potential effects of COVID-related spending, de Soyres, Santacreu, and Young (2023) and Jordà and Nechio (2023) find that countries with more COVID-related fiscal spending exhibit higher post-COVID inflation, and Griffin, Kruger, and Mahajan (2024) examine whether the geographically concentrated patterns of suspicious PPP lending found in this paper are associated with house price growth. Bartik et al. (2023) study whether the use of private banks to disburse PPP funds early in the pandemic was optimal and Aman-Rana, Gingerich, and Sukhtankar (2023) find that heightened documentation requirement for 2021 PPP loans of more than $150k reduced fraud without imposing an undue administrative burden.

been related and exploited.

These findings collectively highlight that in today's increasingly digital and interconnected world, the tools and justification for looting government programs can spread rapidly through social networks. Faster detection, prosecution, and a more proactive administrative approach in government spending programs may be warranted to prevent the spread of new fraudulent schemes. Enforcement might be more effective if it is aware of emerging fraud trends from social media and targets individuals who are nexus points in spreading fraud. Finally, the relation that we document between social media activity related to suspicious financial activity in 2022–23 and activity related to pandemic fraud highlights the potential for persistent future costs of unprosecuted fraud.

1 Background and Data

1.1 Background

We will likely never know the precise magnitudes of pandemic fraud, but estimates from different sources are providing growing information regarding fraud in the PPP, EIDL, and unemployment insurance programs.

The Economic Injury Disaster Loans (EIDL) program is a longstanding program administered by the SBA to provide low-interest rate loans to homeowners and businesses impacted by natural disasters such as hurricanes and floods. In April 2020, the EIDL program was dramatically expanded to allow small businesses to apply for loans to assist with COVID-related economic injuries. In addition to expanding access to loans, Congress also created the EIDL Advance program through which businesses could receive cash infusions of $1,000 per employee, up to $10,000, that did not need to be repaid. The more traditional loan program provided loans for up to $2 million and the two programs totaled more than $400 billion in loans and grants. Because the EIDL was administered directly by the SBA, they have direct access to all data on these loans and grants. A June 27, 2023 report by the Office of the Inspector General (OIG) of the SBA indicates that potentially fraudulent EIDL loans totaled $136 billion, which represents 33% of total disbursed funds.[7]

[7] See report here. The types of fraud indicators used by the OIG for the EIDL program include $55.7 billion based on common or suspicious IP addresses; $34.2 billion due to various hold codes placed on loans by the SBA due to the loan being flagged; $20.5 billion based on duplicated or invalid Employer Identification Numbers (EINs); $31.7 billion based on bank accounts receiving multiple loans or individuals changing their bank account from the bank account listed on their application; $5 billion to sole proprietors or independent contractors without EINs; and the rest due to other indicators such as hotline complaints and suspicious phone numbers, physical addresses, and email addresses.

The Paycheck Protection Program (PPP) targeted small businesses with forgivable loans administered through financial intermediaries totaling $793 billion. Griffin, Kruger, and Mahajan (2023) estimate that PPP fraud totaled approximately $64.2 billion when using individual loan indicators and $117.3 billion using broader county-level indicators, which is 14.8% of the program. The same June 27, 2023 report by the OIG of the SBA, though not having access to as much data as they did for the EIDL such as IP addresses, separately identified $64 billion in potentially fraudulent PPP loans, which represents 8% of total disbursed funds.

Expanded unemployment insurance (UI) programs during COVID amounted to $872.5 billion, and an audit of the UI programs in four large states by the OIG of the Department of Labor found that 20% of Pandemic Unemployment Assistance (PUA) funds were lost to fraud.[8] Based on suspicious debit card transactions, Khetan et al. (2024) estimate that 11.3% of pandemic unemployment insurance disbursements went to suspicious debit cards, and this fraud was facilitated by lax identity verification.

1.2 PPP Data

The main datasets we gather for our analysis consist of loan-level PPP data, measures of social connectedness between geographic areas, advance- and loan-level data for the Economic Injury Disaster Loan (EIDL) program, county-level data on unemployment insurance claims, social media posts, demographic data, and economic data. We start with loan-level indicators of suspicious PPP loans developed by Griffin, Kruger, and Mahajan (2023). These indicators are based on loan-level PPP data released on January 2, 2022 by the Small Business Administration (SBA) and cover all PPP loans issued from the start of the program on April 3, 2020 through the end of the program on June 30, 2021 that had not been repaid or canceled as of January 2, 2022. At the loan level, the data include business name, address, business type (e.g., corporation, LLC, self-employed, etc.), NAICS code (industry), loan amount, number of employees, date approved, loan draw (i.e., initial, first-draw loan or repeat, second draw loan), and lender for 11,469,801 loans originated by 4,809 different lenders with a total value of $793 billion. The primary suspicious loan indicators are LLCs and corporations with missing or ineligible registrations, multiple loans at a residential address, abnormally high implied compensation relative to industry by CBSA averages,

[8] See report here. An analysis by the US Government Accountability Office (GAO) estimates the amount of fraud in pandemic UI programs at between $100 billion and $135 billion (see report here).

and large inconsistencies (as large as tenfold) between the jobs reported by borrowers on their PPP application and jobs reported to the contemporaneous EIDL Advance program, which had a different incentive structure. See Internet Appendix Section A for additional details regarding these measures. Griffin, Kruger, and Mahajan (2023) extensively validate these measures, including with four secondary measures of fraud and three independent external measures. The findings of Griffin, Kruger, and Mahajan (2023) are also validated by a detailed Congressional investigation of PPP fraud that focused on many of the same lenders flagged by Griffin, Kruger, and Mahajan (2023) (see Congressional report here).

For most of our analysis, we aggregate the loan-level indicators of suspicious PPP lending to the zip code level either as a percentage of PPP loans in the zip code or on a per capita basis. Across zip codes, the weighted average suspicious loan rate is 12.4% with a standard deviation of 9.27 percentage points (ppt), and the average zip code has 0.0354 loans per capita.[9] See Table 1 for additional summary statistics.

1.3 EIDL and Unemployment Data

To the best of our knowledge, similar publicly-available granular measures of fraud do not exist for other COVID relief programs. For the EIDL program, a contemporaneous program by the SBA that provided loans of up to $2 million and advances that did not need to be repaid of up to $10,000, we construct several measures of suspicious EIDL advances based on advance- and loan-level data as of December 2, 2020. We first calculate the percent of EIDL advances without a corresponding loan. We also extend the suspicious indicators developed by Griffin, Kruger, and Mahajan (2023) for PPP loans (i.e., inconsistencies of three or more jobs between the jobs reported by borrowers on their PPP and EIDL applications, nonregistered businesses, and multiple loans at a residential address) to EIDL loans and advances, which are discussed in more detail in Section 3.1. Additionally, to study the role of intermediaries in spreading fraud, we submitted a FOIA request to the SBA for data on which agent (if any) facilitated each EIDL loan (described in Section 5.4).[10]

[9]When calculating means and standard deviations of suspicious loan rates, we weight by the number of PPP loans in the zip code or county. This makes the calculations nationally representative and corresponds to the weighting used in subsequent regression analysis. Griffin, Kruger, and Mahajan (2023) also consider a broader measure of suspicious lending for aggregate fraud estimates that does not pinpoint the precise misreported loan but can be calculated at the county level: the share of first draw business loans that exceed the Census establishment counts for a given industry in a particular county. For this potentially more comprehensive measure of potential misreporting, excess share, which we examine for robustness, the weighted average suspicious loan rate across counties is 21.9% with a standard deviation of 13.4 ppt.

[10]Agents could also assist PPP borrowers, but the SBA indicated that they do not have data on this, potentially because PPP loans were facilitated by financial intermediaries.

To determine abnormal unemployment insurance claims, we collect data on initial unemployment insurance claims at the county-month level by visiting each state's Department of Labor (or equivalent) website. For states that do not provide the data on their website, we submitted FOIA requests for the data. Overall, we collect data for 33 states, representing 73% of the U.S. population. Weighted by labor force size, the average county had 42.4 initial unemployment claims per 100 individuals in the labor force between March and December of 2020, with a standard deviation of 22.6 ppt.[11]

1.4 Social Connections Data

For social connections between zip codes, we use data on the strength of Facebook connections between pairs of zip codes based on data from Bailey et al. (2018a, 2020). The strength of these connections is defined as the number of Facebook friendships between users living in zip code i and users living in zip code j normalized by the product of users in zip code i and users in zip code j. Normalizing by the number of users ensures that connection strength measures the intensity of social connections between zip codes independent from the zip codes' intensity of Facebook usage. The data measure connection strength as of 2021, but connections are stable over time (Humanitarian Data Exchange, 2021; Kuchler et al., 2022; Bailey et al., 2021). Because of this stability, connection strength reflects general social connections between zip codes as opposed to anything specific to 2021. As a result, social connections between zip codes should be equally relevant in 2020, and reverse causality from connections formed during the pandemic is unlikely. For each zip code, we use the strength of social connections to other zip codes as weights to calculated the weighted average PPP fraud per capita in socially connected zip codes, which we refer to as social proximity to fraud.[12] We also construct an analogous measure for each geographic zip code's physical proximity to fraud based on the inverse physical distance between geographic areas. We describe the calculations and interpretation of these variables in more detail in Section 4.

[11] To check the accuracy of this data, we aggregate the county-level data to the state-level and compare it to the data released by the Department of Labor (DOL) on initial UI claims at the state-level. The two series match well (see Figure IA.3, Panel B). Per the DOL data, there were 43.3 initial UI claims per 100 individuals in the labor force between March and December 2020 in the 33 states used in our analysis. In Section 3.2, we consider potentially suspicious patterns in the timing and geographic distribution of these claims.

[12] While our main focus is on zip code-level analysis within counties, we also consider county-level social proximity to fraud in robustness tests and when data is only available at the county level. County-level social proximity to fraud is calculated in the same manner.

1.5 Demographic and Economic Data

Demographic data at the zip code and county levels is from the US Census and IRS Statistics of Income. We also consider county-level cultural variables (Parsons, Sulaeman, and Titman, 2018; Grullon, Kanatas, and Weston, 2010; Griffin, Kruger, and Maturana, 2019), FinCEN Suspicious Activity Reports, and identity theft reports from the FTC Consumer Sentinel Network, and zip code-level social capital measures from Chetty et al. (2022a,b). Distances between zip codes and counties are from the NBER. County-level data on employment, spending, and small business revenue during the pandemic are from the Economic Tracker by Opportunity Insights (described in Chetty et al. (2023)). Data on Google search activity is from Google Trends.

1.6 Social Media Activity Data

For social media activity related to pandemic relief programs, we search various social media platforms for discussions regarding PPP, EIDL, unemployment insurance, and related terms. For a large social media platform, we collect data on the membership and posts for 136 groups discussing these topics. Additionally, we collect membership data for 127 groups discussing suspicious financial activity in 2022–23 and 32 groups discussing the Employee Retention Tax Credit (ERC) on the same social media platform.

2 The Geography of Suspicious Lending

If fraud is primarily driven by the idiosyncratic decisions of individual recipients, then one might expect suspicious lending patterns to be distributed equally around the country. Panel A of Figure 1 plots the percent of PPP loans with at least one indicator of suspicious lending in each county across the country. The graph shows significant geographic variation in suspicious lending rates. Areas with a particularly high percentage of flagged loans cluster near New Orleans, Atlanta, and surrounding areas in Louisiana, Mississippi, and Georgia. Chicago and parts of South Carolina also exhibit elevated levels. Many counties in these areas have suspicious lending rates in excess of 25% whereas large parts of the country have suspicious loan rates under 10%. The geographic pattern is somewhat regional, but there are also pockets with elevated rates scattered across the country, including parts of California, Nevada, New Mexico, Arizona, Utah, Texas, Arkansas, most of the southern states, and in northern states such as Illinois, Indiana, Michigan, and Ohio. The pockets appear to cluster in larger cities, but not always. For example, Texas has elevated rates in the mid-size counties around Temple (Central Texas) and Beaumont (East Texas), as well as in the Houston and Dallas-Fort Worth areas, but low rates in San Antonio and Austin. There are also significant differences across large U.S. cities. For example, Cook County (Chicago) has a suspicious loan rate of 31.7% compared to suspicious loan rates of 8.8% in New York County and 6.1% in Los Angeles County.

Does suspicious lending vary even within counties? In Panel B of Figure 1, we examine the relation between suspicious loan rates across counties and zip codes. Each dot represents a zip code, and the size of the dot represents the number of PPP loans in the zip code. The horizontal axis plots the percent of loans flagged at the county level, and the vertical axis plots the percent of loans flagged at the zip code level. There is significant variation across zip codes within counties, as evident from the vertical spread, with flagged loan rates varying from 5% to 35% or more in many counties.

Panel C of Figure 1 shows the distribution of suspicious lending rates over time by county, separately for loans originated by traditional and FinTech lenders. Consistent with the findings of Griffin, Kruger, and Mahajan (2023), suspicious lending rates grew significantly over time, especially for FinTech lenders. In April 2020, 6.1% of loans were flagged as suspicious in the median county for traditional lenders and 6.9% of loans were flagged for FinTech lenders.[13] By May 2021, the median suspicious loan rate grew to 10.0% for traditional lenders and 28.7% for FinTech lenders. Dispersion of suspicious lending rates across counties grew even faster. In April 2020, the range of suspicious loan rates, defined based on the fifth to the ninety-fifth percentile, was 4.0% to 13.3% for FinTech loans. By June 2020, this range increased to 5.7% to 34.3%. The range of FinTech fraud rates continued to increase, particularly during round three of the PPP in 2021, and by May 2021, the range of suspicious lending rates across counties for FinTech lenders was 9.0% to 48.2%. The dots show suspicious lending rates for several large counties that stand out in Panel A of Figure 1. Atlanta (Fulton County) and New Orleans (Orleans Parish) had above-median suspicious lending rates for FinTech loans in April 2020, and these rates continued to grow throughout the program. By contrast, suspicious lending rates in New York County and Los Angeles County grew throughout the program but consistently remained well below the median rate. Chicago (Cook

[13] Percentiles are weighted based on the number of traditional and FinTech loans in each county during the month to make the plot more nationally representative.

County) exhibits a different pattern with an initial suspicious loan rate in April 2020 of 7.7% for FinTech loans, which is similar to the median rate, but by July 2020, its FinTech suspicious lending rate increased to 27.9%. From February to May 2021, Cook County's FinTech suspicious lending rate was close to 50%, which is among the highest rates in the country. Chicago, Atlanta, and New Orleans also had elevated suspicious lending rates for traditional loans, but these rates were significantly lower.[14]

The geographic variance in Figure 1 suggests that fraud is influenced by more than just the isolated decisions of individual PPP recipients. In the next section, we examine how PPP fraud relates to fraud in other pandemic relief programs, and then we ask what explains the stark geographic variation in fraud rates.

3 Connections to Fraud in Other Relief Programs

3.1 Is Suspicious PPP Lending Related to EIDL Fraud?

Businesses were eligible to receive EIDL Advance grants whether or not they accepted or were approved for the loan component of the EIDL program. As a result, EIDL advances were frequently referred to as "free money" in social media posts (see Section 5.1).

The per employee structure of EIDL Advance grants created an incentive to inflate the number of employees reported. Griffin, Kruger, and Mahajan (2023) find that EIDL Advance employee inflation was common, with many instances of recipients reporting ten employees on EIDL Advance applications (thereby maxing out the $10,000 available from the EIDL Advance program) despite only reporting one employee on their PPP application.[15] The incentive for fraud was potentially higher for the advance component of the EIDL than for the loan component because advances were dispersed more quickly and did not need to be repaid. Thus, EIDL advances without a corresponding EIDL loan may be more likely to be suspicious. Consistent with this possibility, 7.6% of advance recipients without a corresponding EIDL loan have ten or more jobs implied by their advance amount and only one job reported on their PPP application, as compared to 0.9% for advance recipients that have a corresponding EIDL loan (as shown in Figure IA.2, Panel A).

[14] Figure IA.1, Panel A shows the distribution of suspicious lending rates over time by zip code. The dispersion of suspicious lending rates across zip codes is even larger for FinTech loans. Figure IA.1, Panel B examines the growth and persistence of suspicious lending rates between 2020 and 2021 across zip codes.

[15] 8.8% of PPP recipients with a matched EIDL advance have at least three more employees implied by their EIDL advance amount than claimed on their PPP application. 43.4% of these recipients have at least 10 employees implied by their EIDL advance amount while they only claimed one employee on their PPP application.

Further, advances without a corresponding loan are also more likely to be flagged as suspicious when suspicious loan indicators developed by Griffin, Kruger, and Mahajan (2023) are applied to the EIDL program (see Table IA.I, Panel A).

Panel A of Figure 2 shows a scatter plot with the percentage of PPP loans flagged as potentially fraudulent in the zip code on the horizontal axis and the percentage of EIDL advances without an EIDL loan in the zip code on the vertical axis. The size of the dots represents the number of EIDL advances in the zip code, and each dot is colored based on the percent of EIDL advances recipients reporting at least three more jobs than they reported on their PPP application. The PPP flag on the horizontal axis is calculated without using the EIDL > PPP jobs flag in order to avoid any mechanical relation.[16] Two patterns are readily apparent. First, zip codes with high PPP flag rates frequently also have high rates of excess EIDL advances. In zip codes with low rates of PPP fraud, around 50% of EIDL advances have no corresponding EIDL loan. In zip codes with the highest rates of PPP fraud, this grows to 80%. Second, the amount of job inflation in the EIDL Advance program across zip codes is increasing with both the amount of PPP fraud and excess EIDL advances. One question is the extent to which this relation between suspicious lending across programs could be due to regional patterns. The same strong relations are also present with county fixed effects, after controlling for demographics, and with alternative measures of suspicious loans and advances based on applying the primary suspicious loan indicators developed by Griffin, Kruger, and Mahajan (2023) to EIDL loans and advances (see Figure IA.2, Panel B and Table IA.II).

3.2 Is Suspicious PPP Lending Related to Excess UI Claims?

Recall that unemployment claims are only available aggregated to the county-month level. Since unemployment rates vary dramatically across states and time, we regress monthly county-level initial unemployment claims per individual in the labor force on county-level PPP fraud per capita, controlling for overall PPP loans per capita, state fixed effects, and state fixed effects interacted with demographics.[17] The regression equation being estimated for each month is:

[16] This means that it is calculated based on flagging multiple loans at a non-business address, firms with invalid or expired business registries, and firms with abnormally high (three times or more) compensation per employee compared to other jobs in the same industry and CBSA.

[17] The regressions are weighted by the number of individuals in the labor force as of December 2019. The demographic control variables are the percentage of adults with at least a bachelor's degree, median income, pre-pandemic unemployment rate, population density, percentage non-white, and poverty rate.

$$UIClaimsDividedbyLaborForce_{i,t} = \beta_t FlaggedPerCapita_i + \gamma_t LoansPerCapita_i + State_{s(i)} + State_{s(i)} \times Demographics_i + \epsilon_i$$

where i is a county, $s(i)$ is the state that county i is in, and t is a month. Standard errors are clustered by state. The coefficient of interest is β_t, which estimates the effect of PPP fraud in a zip code on monthly initial unemployment insurance claims in month t.[18]

The estimated β_t coefficient for each month is plotted in Panel B of Figure 2. During the pre-COVID period spanning January 2019 to February 2020, all monthly coefficients are close to zero and statistically insignificant.[19] This flat pre-trend indicates that the geography of PPP fraud is not correlated with pre-existing unemployment trends. In March 2020, the relation changes, and counties with high PPP fraud begin exhibiting more initial unemployment insurance claims. The β_t coefficient peaks in April 2020 with a coefficient of 1.0, indicating that a one standard deviation increase in PPP fraud is associated with one additional unemployment insurance claim per 100 individuals in the labor force. This relation remains elevated throughout the remainder of 2020 and then returns to zero. The timing and temporary nature of the shock are consistent with excess unemployment insurance claims in the same counties that exhibited high PPP fraud rates. And the magnitude is large. Adding up the coefficients for March to December of 2020 indicates that counties with one standard deviation more PPP fraud had an additional 4.3 initial unemployment claims per 100 individuals in the labor force, which is economically significant compared to mean unemployment claims of 42.2 per 100 individuals in the labor force during this time period.[20] In contrast to the strong relation between UI claims and fraudulent PPP loans, we do not find any evidence that areas with more PPP lending in general have abnormally more UI claims (see Figure IA.3, Panel C).

An alternative explanation for this result is that PPP fraud correlates with actual unemployment shocks. The smaller plots in Panel B of Figure 2 address this possibility by estimating the relations between PPP fraud and employment based on data from the Economic Tracker by Opportunity Insights (described in Chetty et al. (2023)). The relations with the level and change of employment

[18]Throughout the paper, standard errors for regressions estimated using county-level data are clustered by state.

[19]We observe data for 19 states in 2019, 26 states by January 2020, and 33 by March 2020.

[20]This amount may contain double-counting of individuals due to multiple unemployment spells during this period or for other state-specific reporting reasons (which the state fixed effects should account for). Nonetheless, the claims data matches state-level data reported to the Department of Labor, as discussed in Section 1.

are consistently small and statistically insignificant. There is also no relation between PPP fraud and other economic indicators such as spending and small business revenue (see Figure IA.3, Panel D). Regression analysis grouping March to December 2020 together yields equivalent results for unemployment insurance claims and no relation between PPP fraud and other economic indicators (see Table IA.III).[21]

Overall, the evidence in this section suggests that the same areas with heightened levels of PPP fraud also exhibit high levels of suspicious EIDL advances and unemployment insurance claims.

4 Social Connections and Suspicious Lending Patterns

Can social connections between geographic areas explain the geographic clustering documented in Section 2? Following Kuchler et al. (2022) and Kuchler, Russel, and Stroebel (2022), we use the strength of connections between Facebook users in a given location (e.g., county or zip code) and Facebook users in other locations as weights to calculate weighted average suspicious lending rates in socially connected areas, which we refer to as social proximity to suspicious lending. This measure captures the extent to which a given geographic area is exposed to suspicious lending through its social connections to other areas across the country and is independent of the zip code's overall social connectivity and social media usage. To differentiate social proximity from geographic proximity, we also construct a physical proximity measure using inverse distances between geographic areas as weights following Kuchler et al. (2022) and Kuchler, Russel, and Stroebel (2022). Specifically, social and physical proximity to fraud are defined as:

$$SocialProximity_i = \frac{\sum_{j \neq i} SCI_{i,j} \times FlaggedPerCapita_j}{\sum_{j \neq i} SCI_{i,j}}$$

$$PhysicalProximity_i = \frac{\sum_{j \neq i} (1/Distance_{i,j}) \times FlaggedPerCapita_j}{\sum_{j \neq i} (1/Distance_{i,j})}$$

where i and j are two geographic areas (i.e, zip codes or counties); $SCI_{i,j}$ is the social connectedness index between i and j; $Distance_{i,j}$ is the physical distance between i and j, and $FlaggedPerCapita_j$ is the ratio of the number of flagged loans in j to the population of j. Further, the social connectedness index (SCI) is from Bailey et al. (2020) and is calculated as $SCI_{i,j} = \frac{Connections_{i,j}}{Users_i \times Users_j}$ where $Connections_{i,j}$ is the total number of Facebook friendship links between Facebook users living in

[21] Figure IA.3, Panel E shows scatterplots of the relationship between total initial claims from March to December 2020 per individual in the labor force and PPP fraud rates.

i and Facebook users living in j and $Users_i$ is the number of Facebook users in i. In our baseline analysis, we use all zip codes and counties when calculating social and physical proximity to fraud. Results are also robust to calculating social proximity to fraud based only on distant connections such as in other CBSAs or at least of 100, 250, or 500 miles away.

To visualize the social connectedness data, Figure 3 plots network connections between zip codes in three high-fraud Combined Statistical Areas (CSAs): Chicago, Atlanta, and New Orleans. The figure excludes within-CSA connections and focuses exclusively on connections between zip codes across CSAs. Each node is a zip code, and zip codes are sorted into deciles within each CSA based on their PPP fraud rates. The color of each node represents the percentage of PPP loans in the zip code that is flagged as suspicious, with higher fraud rates plotted as darker red. The width of edges between nodes represents the strength of social connections between the nodes (zip codes) being connected. The top decile of fraudulent zip codes in Chicago on the left has strong social connections to the first through fifth highest deciles of fraudulent zip codes in Atlanta. The zip codes with more fraud in New Orleans on the bottom left have strong connections to the highest fraud zip codes in Atlanta. Some of the medium fraud zip codes in New Orleans also have strong social connections to the higher fraud zip codes in Atlanta and Chicago. Overall, high fraud zip codes have a higher tendency to be connected to one another across cities.

4.1 Is Social or Physical Proximity More Important?

Panel A of Figure 4 plots the relation between social proximity to suspicious lending and physical proximity to suspicious lending across zip codes. Each dot is a zip code, and the size of each dot represents the number of loans in the zip code. Social and physical proximity are clearly related, but they also have independent variation, as shown by the vertical dispersion in social proximity to suspicious lending for zip codes with the same physical proximity to suspicious lending. The color of the dots is based on the number of flagged PPP loans per capita in each zip code, with the clear pattern that zip codes with high social proximity to suspicious lending have higher fraud intensity even for zip codes with similar physical proximity to suspicious lending. In other words, focusing on a vertical slice that has a similar level of physical proximity to suspicious lending, the level of flagged per capita (shown by the color of each dot) tends to increase when moving upwards from a low to high level of social proximity to suspicious lending. On the other hand, focusing on a horizontal slice that has a similar level of social proximity to suspicious lending, the level of flagged per capita increases only weakly when moving rightward from a low to high level of physical proximity to suspicious lending.

To more formally examine the relation between social proximity to suspicious lending and suspicious lending intensity, we estimate zip code-level regressions of the following form:

$$\begin{aligned} FlaggedPerCapita_i = {} & \beta SocialProximitytoFraud_i + \gamma PhysicalProximitytoFraud_i \\ & + \theta Controls_i + County_{c(i)} + \epsilon_i \end{aligned}$$

where i is a zip code and $c(i)$ is the county that zip code i is located in; $FlaggedPerCapita_i$ is the ratio of flagged loans in zip code i to its population; $SocialProximitytoFraud_i$ and $PhysicalProximitytoFraud_i$ are social proximity and physical proximity to fraud for zip code i; $County_{c(i)}$ are county fixed effects and $Controls_i$ are demographic control variables.[22] Variables are standardized for ease of interpretation.

The results are shown in Panel A of Table 2.[23] Columns (1) and (2) show that the within-county relation between fraud and social proximity to fraud is similar and highly significant with and without control variables. The coefficient of 1.071 in column (2) indicates that a one standard deviation increase in social proximity to suspicious lending is associated with a 1.071 standard deviation increase in suspicious PPP loans per capita. Column (3) shows that physical proximity to fraud also predicts suspicious PPP lending per capita at the zip code level, but when they are included together in column (4), only social proximity is positive. After controlling for social proximity to fraud, physical proximity to fraud has a slightly negative (though economically small) relation to zip-code-level PPP fraud. Panel B of Figure 4 visualizes the regression from column (4), with the left (right) subpanel showing the within-county relationship between flagged per capita and social (physical) proximity after controlling for physical (social) proximity to suspicious lending

[22] The regressions are weighted by the number of loans in each zip code. Throughout the paper, standard errors for regressions estimated based on zip code-level data are clustered by county. The control variables are population density, percentage non-white, average income, poverty rate, pre-pandemic unemployment, percentage with college education, shares of the friends of Facebook users in the zip code who live within 50 and 150 miles of them, and the FinTech market share of PPP loans in the zip code.

[23] Table IA.IV shows regressions based on the percentage of PPP loans that are suspicious and finds equivalent results, consistent with social transmission of fraud as opposed to demand for PPP loans more generally. Similarly, Table IA.V estimates regressions based on overall PPP lending per capita and non-suspicious lending per capita with much lower results. Table IA.VI shows additional specifications, such as excluding the county fixed effects. Table IA.VII adds more detailed control variables for individual racial groups. Results are also similar when regressions are estimated at the loan level with additional control variables for the number of jobs reported, loan size, lender fixed effects, business type fixed effects, and industry × CBSA fixed effects (see Table IA.VIII).

and the controls described above. Using a broader measure of suspicious lending at the county level, the share of excess loans (described in the Section 1), we find that there are more excess loans in counties that are socially connected to other counties with high excess loan rates, and physical proximity to excess loans has limited predictive power (see Table IA.IX). The main takeaway from Panel A of Table 2 is that social proximity is a robust and much stronger predictor of suspicious lending than physical proximity.

The baseline results in Table 2 focus on Griffin, Kruger, and Mahajan's 2023 combined primary fraud indicator. For robustness, we examine each primary fraud indicator separately and find consistent results (Table IA.X). This helps mitigate potential concerns about non-classical measurement error or omitted variables with respect to any individual measure. We also consider social connectedness data at the county level instead of at the zip code level (Table IA.XI). When measures are constructed separately for FinTech and traditional loans, we find strong results for FinTech but not for traditional loans, which is consistent with Griffin, Kruger, and Mahajan's 2023 finding that fraud is concentrated in FinTech loans (Figure IA.4). Finally, to ensure that our results are not specific to social connections as measured by the Facebook social graph, we construct an alternative measure of social proximity based on taxi and rideshare data from Chicago and find very similar results (Figure IA.5 and Table IA.XII).

Our social connection analysis is focused on PPP fraud mainly due to the richness of the PPP data. Conceptually, we would expect the geographic clustering of fraud in other pandemic relief programs to also correlate with social connections. We examine this for the EIDL program since we have similarly granular data for it, and we find that social proximity to suspicious EIDL loans and advances is a robust predictor of suspicious lending rates and drives out the effects of physical proximity (see Table IA.XIII).

4.2 Are the Effects of Social Proximity Concentrated in Particular Areas?

To assess whether the relation between social proximity and suspicious lending is concentrated in zip codes with particular demographics, we add interactions between social/physical proximity to fraud and indicator variables for whether a zip code is above or below the median of different demographic characteristics (also controlling for the demographic indicator variables themselves). Specifically, we estimate regressions of the form:

$$\begin{aligned} FlaggedPerCapita_i = {} & \beta_{below}(Proximity_i \times 1(Demographic_i < DemographicMedian)) \\ & + \beta_{above}(Proximity_i \times 1(Demographic_i \geq DemographicMedian)) \\ & + \gamma 1(Demographic_i > DemographicMedian) \\ & + \theta Controls_i + County_{c(i)} + \epsilon_i \end{aligned}$$

where i is a zip code and $c(i)$ is the county that zip code i is located in; $FlaggedPerCapita_i$ is the number of flagged loans in i divided by the population of i; $Proximity_i$ is either social or physical proximity to suspicious lending for i; $Demographic_i$ is the value of the demographic that the split is being performed with respect to in i; $DemographicMedian$ is the median value of the demographic; $County_{c(i)}$ are county fixed effects and $Controls_i$ are the same zip code-level control variables as previous regressions. The coefficients of interest are β_{below} and β_{above}, which estimates the effect of social/physical proximity on fraud rate in zip codes with below- and above-median values of the demographic, respectively.

Figure 5 plots the results, with separate regressions for social proximity to suspicious lending (Panel A) and physical proximity to suspicious lending (Panel B). The first column of Panel A plots coefficients for the effect of social proximity to suspicious lending on suspicious PPP lending rates in low- and high-income zip codes. Results are almost identical and remain large and highly significant for both subsets of zip codes. The same is true for zip code subsets based on poverty rates, population density, minority population share, educational attainment, and pre-pandemic employment. For example, the effect of a one standard increase in social proximity on fraud intensity for zip codes with below median percentage non-white is 0.96 versus 1.07 in zip codes with above median percentage non-white, a difference that is not economically nor statistically significant.[24] Estimates are also similar when we split zip codes based on their population share of specific racial groups (see Figure IA.6). Almost identical coefficients across zip codes with different demographics point to a broad-based effect of social proximity as opposed to an effect that is concentrated in a particular demographic group. Panel B of Figure 5 shows the equivalent results for physical proximity. Consistent with previous results, physical proximity to suspicious lending has a weak relation with suspicious lending rates. However, as with social proximity, the magnitude

[24] All demographic variables are from the US Census American Community Survey. The poverty rate is the percentage of households with income below the poverty threshold, which varies based on family size and composition. Educational attainment is the percentage of adults with at least a bachelor's degree. Minority population share is the percentage of non-white individuals.

of this relation is nearly identical across all subsets of zip codes. Consistent results across zip codes with diverse demographics are reassuring and may alleviate (but not eliminate) some concerns about omitted variables, which we explore in more detail next.

4.3 Are Social Connection Results Due to Homophily or Omitted Variables?

A potential concern with the results in Figure 4 and Panel A of Table 2 is that they could be influenced by homophily or omitted variables if socially connected zip codes are similar along dimensions that relate to PPP fraud. For example, suppose that poorer areas or areas with larger minority populations have higher fraud rates and are also more likely to be socially connected to one another. If this is the case, social proximity to fraud could be correlated with an omitted variable related to poverty, education, or race, which could bias the coefficient estimates.[25] The inclusion of these detailed control variables including population density, percentage non-white, household income, poverty rate, pre-pandemic unemployment, and educational attainment and the examination of the impact of culture in Section 4.4 help mitigate this concern because any homophily would have to be along unobserved dimensions. Our results are also unaffected by adding more detailed control variables for individual racial groups (see Table IA.VII). Adding control variables essentially controls for homophily along observed dimensions. Following the logic of Oster (2019), the nearly identical results with and without control variables in columns (1) and (2) suggest that for omitted variables to significantly bias the coefficient, unobserved variables would have to be much more powerful or much more correlated with PPP fraud than the observed controls.[26]

As discussed above, the primary concern is that demographically similar areas may have similar rates of fraud and also be more socially connected to one another. To more directly test this specific explanation, we construct a measure of demographic proximity between zip codes. Demographic proximity to suspicious lending is calculated equivalently to social proximity to suspicious lending except that the weights are based on average similarity along six demographic dimensions between

[25] Along these lines, significant coefficients on the demographic control variables indicate correlations between fraud intensity and demographic characteristics (see Table IA.VI, Panel B).

[26] Using the approach advocated by Oster (2019), if unobserved homophily and other omitted variables have the same impact on the social proximity coefficient as observed demographics (equal selection assumption) and could potentially increase the R^2 of the regression from 0.796 to 0.90, the coefficient on social proximity in column (2) would only decrease from 1.071 to 0.895 $[0.895 = 1.071 - (0.90 - 0.796) \times (1.110 - 1.071)/(0.796 - 0.773)]$. Even under a more extreme assumption that omitted variables could increase the R^2 to 1.0, the coefficient would be 0.725, which is still economically large. Relaxing the equal selection assumption and maintaining the assumption that omitted variables could increase the R^2 to 0.90 (1.0), the unobserved homophily and other omitted variables would need to have approximately 6 (3) times the proportional impact of the observed control variables to decrease the coefficient to zero.

each zip code pair.[27] Column (1) of Panel B of Table 2 estimates a regression of flagged PPP loans per capita on demographic proximity to suspicious lending with county fixed effects.[28] The coefficient of 0.360 is statistically significant and economically large, which indicates a general tendency for zip codes with similar demographics to have similar suspicious lending per capita. In column (2), we include both social proximity and demographic proximity to test which is more important. In this specification, demographic proximity has no relation to suspicious lending after controlling for social proximity, and the coefficient on social proximity is almost identical to columns (1) and (2) of Panel A. We also consider versions of the regressions in columns (1) and (2) with separate demographic proximity to suspicious lending measures for each demographic variable with the same basic result (see Table IA.XIV). The consistent implication of these results is that social proximity to fraud is distinct from homophily at least along observed dimensions, and demographic similarity appears to be much less important than social connections.

In columns (3) and (4) of Table 2, Panel B we estimate an IV version of the regression in column (2) of Panel A using social proximity based on far distant zip codes as an instrument for overall social proximity to suspicious lending. Restricting to distant social connections has the benefit of avoiding any omitted variables related to local shocks such as socially connected individuals having shared in-person exposure to fraud (e.g., a local advertising campaign).[29] In column (3), we restrict to zip codes that are over 100 miles away, and in column (4), we restrict to zip codes that are at least 500 miles away. The IV results in columns (3) and (4) of Panel B are similar to the OLS results in Panel A, and if anything, are slightly higher. Results are similar with alternative distance thresholds, and including multiple instruments based on different distance ranges passes the overidentification J-test (see Table IA.XVI, Panel A). Overall, the IV results reinforce the OLS results and highlight that social connections have a large impact on fraud even when restricted to zip codes that are hundreds of miles apart.[30]

[27] The demographics used are population density, percentage non-white, household income, poverty rate, pre-pandemic unemployment, and educational attainment. Similarity between zip code pairs along each of these demographics is defined as $1-|PercentileRank(Demographic_i)-PercentileRank(Demographic_j)|$ where $PercentileRank(.)$ is a percentile rank between 0 and 1, i and j are two zip codes, and $|.|$ is the absolute value operator. Demographic proximity to fraud is correlated with social proximity to fraud, but only moderately (see Table IA.XV).

[28] Control variables are omitted from this regression because controlling for zip code demographics has the effect of controlling for homophily along observed dimensions.

[29] This IV follows Bailey et al. (2018b,c), which use the house price experiences of an individual's distant friends as an instrument for the house price experiences of all friends. The intuition is that the instrument only uses social connections that are distant and unlikely to be due to local shocks or physical contact. See Table IA.XVII for the first stage of the IV regressions.

[30] Table IA.XVI, Panel B and Figure IA.7 show the reduced form of these IV regressions. Table IA.XVIII estimates loan-level versions of these IV regressions and finds equivalent results.

A final approach for assessing the likelihood of results being driven by homophily is to examine evidence of specific mechanisms for the transmission of fraud through social connections. This is the focus of Section 5, which traces individual-level fraud to specific social media groups with prominent fraud discussions and also shows that social connections predict specific decisions such as what lender and agent to use. These patterns are what we would expect from social transmission of fraud and would be difficult to explain if social connections merely proxied for demographic similarity.

4.4 Is Fraud Explained by Local Culture?

What role does culture play in explaining fraud? Do culture, historic rates of identity theft, or historic levels of suspicious financial activity play a role in spreading suspicious lending? Other types of fraud exhibit regional concentration (Parsons, Sulaeman, and Titman, 2018), and investigators have found identity theft to be common in unemployment insurance fraud during the pandemic (e.g., ProPublica 2021). The same could be true for PPP fraud. The regressions in Table 2 control for demographics and include county fixed effects to alleviate this concern as much as possible. To assess the role of cultural factors, historic rates of identity theft, and historic levels of suspicious financial activity, we replicate the same analysis at the county level (with state fixed effects) to examine county-level demographics, cultural measures, and past rates of identity theft and suspicious financial activity. The cultural variables we examine are public corruption convictions, religious affiliation, and Ashley Madison usage, a proxy for marital infidelity, which have all been shown to predict financial fraud in other studies (Parsons, Sulaeman, and Titman, 2018; Grullon, Kanatas, and Weston, 2010; Griffin, Kruger, and Maturana, 2019).[31]

Table 3 reports the results. All variables are standardized to have a standard deviation of 1 to aid in comparison. Columns (1) and (2) show that at the county level social proximity to suspicious lending is strongly related to fraud with and without controlling for physical proximity to suspicious lending. In column (3), we add potential county-level fraud predictors and find that they have limited incremental predictive power. The coefficient for contemptuous identity theft is significant at the 1% level, for past rates of identity theft and religious affiliation at the 5% level, and for contemptuous suspicious activity reports at the 10% level. Using contemporaneous identity theft and SARs in column (4) results in positive coefficients but adding these control variables has

[31] See the table header for a description of each of the cultural variables.

little impact on the coefficient for social proximity to suspicious lending. Further, the incremental explanatory power of the cultural variables is relatively weak, as evidenced by the small incremental R^2 when the cultural variables are added. Overall, the evidence shows that social proximity is a strong predictor of PPP fraud, and other known predictors of fraud do not explain this relation.

We next examine differences in social capital based on measures developed by Chetty et al. (2022a,b). The regressions in Table 4 jointly estimate the effects of social proximity to suspicious lending and social capital on PPP fraud rates at the zip code level. Social proximity strongly predicts fraud regardless of which social capital variables are included in the regression as control variables. Social capital itself generally has a positive relation with fraud after conditioning on social proximity to suspicious lending. When all social capital measures are included together, the strongest social capital predictor of fraud is the number of civic organizations followed by clustering, which measures the rate at which two friends of a person are also friends with each other. Chetty et al. (2022a,b) find that clustering can reinforce pro-social behavior. Our results suggest that clustering may also support the spread of negative behavior. Economic connectedness, the share of high socioeconomic friends among low socioeconomic individuals, is also weakly positively related to fraud rates. By contrast, volunteering rates, another measure of social capital, are weakly associated with lower levels of PPP fraud. All of the social capital measures have economically small relations to PPP fraud rates compared to social proximity to suspicious lending.[32]

4.5 Do Social Connections Explain the Spread of Suspicious Lending Over Time?

If suspicious PPP lending spread through social connections, one might expect it to spread from areas with high initial suspicious PPP lending rates. To test this, we first identify zip codes in the top ten percent of suspicious lending rates (weighted by PPP lending volume during the first month of the program). We then sort all the remaining zip codes into deciles according to their social proximity to the initial high fraud zip codes.[33]

Panel A of Figure 6 tracks how the rate of suspicious lending changed across the social connection

[32] Table IA.XIX examines the relations between social capital and fraud rates without including social proximity to suspicious lending in the regression. Support ratio and volunteering rate have a negative relation to fraud, while civic organizations are again positively related to fraud.

[33] We measure each zip code's social connectedness with the initial high fraud zip codes in a manner analogous to Hu's (2021) population weighted measure. That is, for each zip code i, $SocialConnectedness_{i,InitialHighFraud} = \sum_{j \in InitialHighFraud} Population_j \times SocialConnectedness_{i,j}$.

deciles, as well as in the initial high fraud zip codes. The bottom row shows the initial high fraud zip codes. By construction, these zip codes had the highest fraud during the first four weeks of the program. Fraud rates in these zip codes grew from 13% during the first month to over 30% in the closing weeks of the program. The remaining zip codes are sorted into deciles based on their social proximity to the initial high fraud zip codes. During the first month of the PPP, fraud rates were below 7% for all deciles. Over time, fraud growth is particularly pronounced in the zip codes with the strongest social connections to the initial high fraud zip codes. By the closing weeks of the program, the top two deciles of highly socially connected zip codes had even higher fraud rates than the initial high fraud zip codes themselves. By contrast, the lowest decile of socially connected zip codes had fraud rates under 16% throughout the program.

We next perform a similar analysis with zip codes matched based on initial fraud rates. For each zip code in the top tercile of social connections to the initial high fraud zip codes, we match it (with replacement) to a zip code in the bottom tercile of social connections that has a similar rate of suspicious lending during the first month of the program. From Panel B of Figure 6, one can see that the resulting matched top and bottom tercile zip codes have, by construction, nearly identical rates of suspicious lending during the first month of the program.[34] However, the rate of suspicious lending grows much more rapidly in rounds 2 and 3 for the top tercile. Thus, despite the matched top and bottom social connection zip codes having identical flagged rates at the start of the PPP, by the end of the program 35% of loans in the top tercile of social connectedness with the initial high fraud zip codes are suspicious while less than 20% are flagged in the bottom tercile. This difference between the suspicious lending rates is highly statistically significant, as shown by the 95% confidence intervals represented by the dashed lines.[35] We find similar results when zip codes are split into social connection terciles and matched on initial fraud rates within counties: suspicious lending rates are similar between the top and bottom tercile in the first month of the program by construction, but suspicious lending grows much faster in the top tercile over time (see Figure IA.8, Panels B and C).

The analysis of how PPP fraud spread has thus far focused on relatively quick spread at the

[34] As a quantification of this, the p-value for the difference between flag rates in the top tercile of zip codes and the matched zip codes in the bottom tercile is 0.67. The rates of suspicious lending over time in each tercile, before matching, and the initial high fraud zip codes are shown in Panel A of Figure IA.8.

[35] Standard errors are clustered by zip code to account for any repetition in the matched bottom tercile zip codes.

weekly frequency. We next turn to how fraud spread from rounds 1 and 2 of the PPP (from April to August 2020) to round 3 (from January to June 2021).[36] Did 2020 PPP fraud spread to connected areas in 2021?[37] For this analysis, we focus on the cross-section of zip codes and regress 2021 fraud rates on 2020 fraud rates in socially and physically proximate zip codes as well as the zip code's own 2020 fraud rate. In particular, we estimate regressions of the form:

$$FlaggedPerCapita2021_i = \beta SocialProximitytoFraud2020_i + \gamma PhysicalProximitytoFraud2020_i + \delta FlaggedPerCapita2020_i + \theta Controls_i + County_{c(i)} + \epsilon_i$$

where i is a zip code and $c(i)$ is the county that zip code i is located in; $FlaggedPerCapita2020_i$ and $FlaggedPerCapita2021_i$ are the ratios of flagged loans in 2020 and 2021 in zip code i to its population, respectively; $SocialProximitytoFraud2020_i$ and $PhysicalProximitytoFraud2020_i$ are social proximity and physical proximity to fraud for zip code i based on loans made in 2020; $County_{c(i)}$ are county fixed effects and $Controls_i$ are the same zip code-level control variables as previous regressions.

Table 5 reports the results. Column (1) examines the effects of social connections using fraud rates based on all PPP loans. A one standard deviation increase in social proximity to 2020 fraud is associated with a 0.885 standard deviation increase in a zip code's 2021 fraud rate. In addition to county fixed effects and other control variables, this regression also controls for each zip code's own 2020 fraud rate. Fraud rates are persistent, resulting in a coefficient of 0.472 on the zip code's 2020 fraud rate, meaning that a one standard deviation increase in a zip code's 2020 fraud rate is associated with a 0.472 standard deviation increase in its 2021 fraud rate. Column (2) considers physical proximity to 2020 fraud using the same framework. The coefficient on physical proximity is small and insignificant, indicating that PPP fraud spread through social connections as opposed to physical proximity. Including social and physical proximity to 2020 fraud together in column (3) results in a nearly identical coefficient for social proximity. Overall, columns (1) to (3) of Table 5 show that fraud spread geographically from 2020 to 2021 at least in part through social connections.[38] Table 5 also shows that the effect is nearly entirely through FinTech loans and there

[36] Round 3 gave borrowers an opportunity to obtain a second loan and also opened up the program to many first-time PPP borrowers.

[37] Figure IA.9 graphically shows that a zip code's social proximity to fraud based on 2020 FinTech loans is predictive of the zip code's 2021 FinTech fraud rate (Panel A) and also of its growth in FinTech lending between 2020 and 2021 (Panel B). A zip code's social proximity to fraud based on 2020 traditional loans is neither predictive of the zip code's 2021 traditional fraud rate nor of its growth in traditional lending between 2020 and 2021.

[38] Table IA.XX, Panel A estimates similar regression at the zip code-month level and shows that even after controlling for

is almost no evidence of social transmission of fraud in traditional loans.

The overall implication is that fraud spread quickly through social connections. The next section examines whether specific information, such as loan features, lender choice, and usage of specific agents, also spread through social connections.

5 Mechanism

How did pandemic fraud spread through social connections? One longstanding challenge to isolating peer and social effects is whether homophily among peers might explain inferences. Manski (1993) overviews this challenge and calls for "richer data" to address the problem. Following this suggestion, we focus on transmission through individual social media groups and the relation between social connections and specific loan characteristics including choice of intermediaries and concentrations of identical loan features. If social connections can explain detailed information such as clusters of identical loans and decisions about specific lenders to use, this is more likely to be explained by transmission of information as opposed to general similarities between areas.

5.1 Social Media Activity

How does information about fraud opportunities spread between borrowers? We examine social media activity related to the PPP and other pandemic relief programs to learn about the nature of the content discussed, the extent of related social media activity, and linkages between social media activity and suspicious PPP loans at the individual level. We find 233 groups on a single large social media platform discussing the PPP and related topics.[39]

The terms used and the nature of the social media activity vary considerably. Some activity appears to be related to legitimate help with filling out forms with group names including "SBA Loan Advisor," "EIDL/PPP FAQs," "PPP Help," and "SBA Loan Guide." However, other posts and groups appear to be mainly focused on opportunities for fraud with frequent discussion of "special sauce" and "methods." Examples of such activity include groups and posts mentioning "Official Docs & CPNS Methods ARE NOT FREE," "SBA PPP Sauce... I can help with any documents you need, statements, invoices, IDs you name it, I got it.," "#PPP scam," and "All Kinds

the zip code's fraud rate in the previous month, zip code fixed effects, month × county fixed effects, month × demographics, month × percentage of FinTech loans, the zip code's social proximity to fraud in the previous month predictive of the zip code's fraud rate in the given month. Table IA.XX, Panel B estimates similar regressions at the county-month level and finds similar results.

[39] We identify these groups by searching on this platform for terms such as PPP, EIDL, SBA, "sauce," and "methods." Figure IA.10 repeats the following analysis using groups found only by searching for PPP, SBA, and their full names. The results are qualitatively the same as those described below.

Of Methods Here, SBA EDD CALI, CREDIT CARD.....ALSO FOR SAUCE." We provide examples of such posts on a number of different social media platforms, including Facebook, Instagram, YouTube, Telegram, Reddit, and TikTok, in Exhibit IA.1. A common theme across many of the most questionable posts and groups is that they advertise multiple programs at once, including PPP, EIDL, and unemployment insurance, potentially indicating linkages between fraud across these programs.

To understand the role of social media in enabling fraudulent activity, we next examine the extent to which individuals in suspicious social media groups actually took out suspicious PPP loans. We start by collecting members from 136 public groups on a single large social media platform.[40] Then we match individuals in the groups to the loan-level PPP data using their name. To reduce the number of false matches, we focus on unique names and include only users that are matched to a single person in the PPP data.[41] Overall, across the 123,263 unique users in the 136 groups, 17,615 (14.3%) are matched to the PPP data.[42]

Figure 7, Panel A plots the PPP flag rates among members of the 72 groups with at least 25 members with matched loans.[43] The rate of flagged loans varies substantially across groups, and the names of many of the groups with higher flag rates contain suspicious words such as "document," "cash," "tap," "cpn," "method," and "sauce."[44] In contrast, many of the groups with below-average flag rates seem to be focused on providing advice and answering questions regarding pandemic relief programs.

To further understand potential differences in discussion content across groups, we collect posts made in the ten groups with the highest and lowest flag rates among the PPP loans received by

[40]Some of the groups identified above are private groups, so we are unable to access them. We choose this platform because we found the most activity discussing pandemic relief programs on it, people usually use their real names on the platform, and we can use information on users' profiles to validate our matching.

[41]We are unable to collect all of the members due to the platform only listing a subset of members and due to technical limits for the larger groups. We only match borrowers who took out loans in their name and not in the name of a business. The matching procedure focuses on unique names conditional on receiving a PPP loan. We allow a user to be matched to at most one loan in each draw of the PPP. Note that this requirement does not exclude loans flagged by the multiple loans flag since such loans are almost always taken out under different borrower names. We manually verify a random sample of 200 matched loans using LexisNexis and find that at least 86% are correct matches.

[42]These 17,615 members received 21,321 PPP loans for a total of $336 million. 76.5% of loans were originated by a FinTech lender.

[43]For ease of viewing, we only show the 10 groups with the highest and lowest flag rates across the loans matched to members of each group. Figure IA.11 shows a version of this figure with all 72 groups. The number of members matched to PPP loans and the total number of members in the group are shown to the right of the bars. The number of users in each group (noted in parentheses) varies substantially from 119 to over 23,000.

[44]The error bars represent 95% confidence intervals with Bonferroni correction for multiple testing. Figure IA.12 shows that the differences in flag rates across groups are large even after controlling for CBSA, business type, loan amount, and jobs reported. In all three specifications shown in Figure IA.12, the group fixed effects are jointly statistically significant.

their members. The most commonly used 200 words in each set of groups are plotted as word clouds in Panel B of Figure 7. The left word cloud shows the most common words in the ten groups with the lowest flag rates, and the right plot shows the most common words in the ten groups with the highest flag rates. Words that are used primarily for potentially nefarious purposes are shown in red.[45] The groups with high PPP flag rates use substantially more nefarious words and also frequently discuss Womply and BlueAcorn, the two lending platforms that are affiliated with six of the PPP lenders that Griffin, Kruger, and Mahajan (2023) find to have among the highest suspicious lending rates.

A key proposition of Sutherland's (1939) differential association theory is that the conditions for crime consist of not only the skills and techniques to commit the crime but also the definitions and rationalizations that normalize crime. While the sheer scope of posts promoting and advertising pandemic fraud may help in its normalization, it is worth highlighting some examples, such as a YouTube music video with over 2 million views that promotes the PPP and includes a man holding signs that state "you too can get a PPP loan" and "it's nobody business if you have a business." Another YouTube music video with over 1.1 million views includes pictures of a PPP application surrounded by cash and lyrics like "I've been running up that bag." A Reddit post argues that "the reality is that a $20k PPP loan is just a drop in the bucket compared to the tax fraud committed by millionaires and large businesses," and another stating, "literally everyone I know done it [sic] and they never got caught, . . . it's literally impossible to prosecute every single person who took one out, they don't have the resources for that."[46]

5.2 Loan Application Details

Social media activity includes frequent references to secret "sauce" and "methods" for obtaining fraudulent loans. To the extent that socially connected fraudulent applicants use and share the same methods, they may have suspiciously similar loan applications. Griffin, Kruger, and Mahajan (2023) find that suspicious loans are frequently associated with clusters of loans highly similar features such as identical loan amounts, identical number of employees, and excessive concentrations in

[45] For example, "method," "sauce," "glitch," and "drop" are used to mean the poster has information on how to access some form of money. "Document," "doc," and "package" frequently refer to the poster selling documents necessary to apply to relief programs. "Inbox," "dm" (short for "direct message"), "pm" ("private message"), "hmu" ("hit me up"), and "hit" are used when the poster wants to share information through more private means. "Upfront," "deal," "PayPal," "Cash App," "Apple Pay," "split," and "free" are used in the context of payment for the "methods" or "sauce." "Chime" and "Varo" are two banks that are commonly recommended for receiving funds due to their ease of use.

[46] See Exhibit IA.1 for documentation of these posts and other examples.

particular industries, often resulting in more PPP loans to an industry in a county than the number of establishments that exist based on Census data.

As extreme examples of loan clustering, Griffin, Kruger, and Mahajan (2023) point to 4,299 first-draw loans for $20,000 originated by Cross River Bank to Illinois businesses in the "Insurance Agencies and Brokerages" industry, almost all of which have exactly one reported employee. Another 938 Cross River first-draw loans for $20,000 to businesses in Illinois, also concentrated in Cook County, cluster in the "All Other Miscellaneous Crop Farming" industry with the majority of recipients reporting exactly eight employees.[47] To understand whether social transmission of information could play a role in such loan clustering, we explore whether counties that are socially connected to Cook County also have large clusters of loans originated by Cross River with the same identical features. For every county, we determine the percentage of Cross River loans in the county that exactly match the loan clusters found in Cook County along all three features (i.e., they are in the same industry, for the same amount, and report the same number of employees). Then, we examine whether this percentage is related to how socially connected the county is to Cook County.

Panel A of Figure 8 shows the results as a binscatter with state fixed effects.[48] The left subpanel considers loans for $20,000 to businesses in the "Insurance Agencies and Brokerages" industry with one employee reported. As a county's social connectedness to Cook County increases, a larger percentage of Cross River's loans in the county exactly match the cluster of loans found in Cook County on all three loan features. Counties in the top quintile of social connectedness to Cook County have 0.43 ppt more of Cross River's loans matching all three loan features compared to counties in the bottom quintile. This is a large effect given an average county (excluding Cook County) has 0.78% of Cross River's loans matching all three of these loan features. It is also worth noting that there are only 392 loans across the entire country that match along all three features and were not originated by Cross River. The percentage of loans originated by other lenders that have features identical to the cluster are plotted in orange (other FinTech lenders) and green (traditional

[47] The examples described by Griffin, Kruger, and Mahajan (2023) are restricted to first-draw loans. We generalize this by including all loans and require that clusters have identical reported employees. Focusing on Cook County, where these loans concentrate, these clusters of identical loans represent 11.0% and 1.3% of all loans originated by Cross River in Cook County, respectively.

[48] The binscatter for each lender or set of lenders is run separately and is weighted by the number of loans by the lender in each county. Results are also robust to excluding all counties in Illinois and controlling for physical distance to Cook County.

lenders). In both cases, the incidence of these loans is almost exactly zero regardless of distance to Cook County. The right subpanel considers loans for $20,000 to business in the "All Other Miscellaneous Crop Farming" industry with eight employees reported and finds analogous results.

Building on these two examples, we identify additional clusters of loans among loans originated by Cross River and the two large online platforms that facilitate the majority of FinTech loans, Womply and BlueAcorn, and test whether clusters of loans with identical loan features tend to concentrate in socially connected counties more generally. To identify clusters of loans, we determine combinations of identical industry, loan amount, and reported employees with at least 100 loans originated by a single lender/platform in a single county.[49] Using these criteria, Cross River has 40 clusters with identical loan features (representing 8.0% of Cross River's overall loans across the country), Womply has 40 clusters (representing 4.1% of loans), and BlueAcorn has 21 clusters (representing 5.1% of loans). Among traditional lenders, there are no clusters of identical loans meeting these criteria, either for individual lenders or for traditional lenders overall. We then identify the county in which the lender has the most loans matching each combination of loan features as the central county for each cluster's network. Finally, we repeat the analysis described above for each cluster of loans and average the results across all clusters found for each lender.

Panel B of Figure 8 shows the results with the left subpanel based on clusters identified for Cross River, the center panel for Womply, and right panel for BlueAcorn. For all three lenders, we find that as a county's social connectedness to central county in the cluster's network increases, a larger percentage of lender's loans in the county exactly match the cluster on all three loan features. Counties in the top quintile of social connectedness to the central county have 0.079, 0.026, and 0.052 ppt more of Cross River's, Womply's, and BlueAcorn's loans matching all three loan features compared to counties in the bottom quintile, respectively. These are large effects given that an average county (excluding the central county) has 0.18%, 0.094%, and 0.23% of loans exactly matching the cluster, respectively for Cross River, Womply, and BlueAcorn.[50] Incidence of loans originated by other lenders that have features identical to the cluster are plotted in orange

[49]We exclude loans where the loan amount is at the maximum allowed per employee to avoid loans that may cluster at this point mechanically. We define industry based on 4-digit NAICS codes and round loan amounts to the nearest $100 to allow for small differences. The Womply and BlueAcorn platforms are defined based on lenders with relationships to these platforms as described in more detail in the next subsection.

[50]As a point of comparison, on average across clusters, the central county has 1.05%, 1.11%, and 0.44% of loans in the cluster, respectively for Cross River, Womply, and BlueAcorn.

(other FinTech lenders) and green (traditional lenders). Loans with features identical to the Cross River clusters are mostly limited to Cross River itself. For BlueAcorn and Womply, other FinTech lenders also exhibit elevated rates of loans with identical features, particularly in counties that are socially connected to the center of the cluster. This is consistent with the same "sauce" and "methods" being promoted for usage across multiple lenders and platforms, particularly for Womply and BlueAcorn, which is what we see in the social media activity described above.

5.3 Specific PPP Lenders

Information sharing through social connections and lender advertising on social media could embed different FinTech lenders within different social networks. In addition to discussing general "methods" and "sauce," many of the previously described social media posts also discuss which FinTech lenders to use for easy approval. In particular, the largest two FinTech platforms for PPP lending, BlueAcorn and Womply, feature prominently within the suspicious PPP groups shown in Figure 7 and these platforms appear to have advertised extensively on social media (see Congressional report here).[51] Additionally, using loan-level data matched to social media users at the individual level, we find that borrowers in social media groups that discuss BlueAcorn and Womply are more likely to use these platforms for their own loans (see Table IA.XXI).

To evaluate whether social networks influenced lender choice more generally, we estimate regressions of the form:

$$\begin{aligned} LoansPerCapitaByLender_i = {} & \beta SocialProximitytoLender_i + \gamma LoansPerCapita_i \\ & + \theta Controls_i + County_{c(i)} + \epsilon_i \end{aligned}$$

where i is a zip code and $c(i)$ is the county that zip code i is located in; $LoansPerCapitaByLender_i$ is the ratio of loans made in zip code i by a given lender (or group of lenders) to population of zip code i; $SocialProximitytoLender_i$ is the social proximity to lending by a given lender (or group of lenders) in other zip codes and is defined in the same way as social proximity to suspicious lending, but with the number of flagged loans per capita in each zip code replaced with the number of loans originated by the lender(s) per capita in the zip code; $LoansPerCapita_i$ is the ratio of loans made in zip code i by any lender to the population of zip code i; $County_{c(i)}$ are county fixed effects and

[51] Examples of advertising by FinTech lenders, in particular Womply and BlueAcorn, regarding the PPP are provided in Exhibit IA.2.

$Controls_i$ are the same zip code-level control variables as previous regressions.

Consistent with social networks influencing lender choice, column (1) of Table 6 shows that a one standard deviation increase in a zip code's social proximity to FinTech lending is associated with a 0.86 standard deviation increase in FinTech loans per capita. Further, the choice of specific lenders also appears to be influenced by social connections for the two largest FinTech platforms (Womply and BlueAcorn) but not for traditional lenders. Specifically, a one standard deviation increase in a zip code's social proximity to Womply (BlueAcorn) lending is associated with a 1.02 (1.09) standard deviation increase in Womply (BlueAcorn) loans per capita (columns (2) and (4)), and these relations persist even after controlling for social proximity to the other FinTech platform (columns (3) and (5)).[52] By contrast, social proximity to Bank of America and JPMorgan Chase has only a small and statistically insignificant relation to loan volume for those lenders. These relations highlight that social connections influenced specific decisions such as what FinTech lender to use, which is not something we would expect if social connections merely proxy for homophily between zip codes.

5.4 Specific Agents in EIDL Lending

An interesting feature of the EIDL application is that it asked for the name and address of anyone who helped the borrower apply and also asked whether any fee was charged or agreed to.[53] In response to our FOIA request, the SBA released data on agents who helped EIDL borrowers with their applications. In particular, for every loan that had any information filled out for the aforementioned question, the data includes an anonymized ID for the agent, the fee charged or agreed to (if any), and the EIDL loan ID. In combination with the loan-level EIDL data that the SBA previously released and loan-level fraud indicators for the EIDL program (see Section 3.1), this allows us to examine the potential network effects of agents in spreading fraud.

We first ask if social networks predict which agent borrowers use. We find that a 1 ppt increase in the relative probability of an individual from zip code i having a friendship with an individual

[52] As discussed by Griffin, Kruger, and Mahajan (2023), Womply worked with five lenders to originate loans: Benworth Capital, Capital Plus Financial, DreamSpring, Fountainhead SBF, and Harvest SBF. BlueAcorn worked with two lenders to originate loans: Capital Plus Financial and Prestamos CDFI. Because Capital Plus worked with both Womply and BlueAcorn, Capital Plus loans are included as lending activity of both platforms for Table 6. Table IA.XXII replicates columns (2) to (5) of Table 6 after excluding Capital with extremely similar results. Figure IA.13 graphically shows the interaction between social proximity to Womply and BlueAcorn lending and their relations with Womply and BlueAcorn loans per capita. Physical proximity to FinTech and specific lenders has limited predictive power, especially once social proximity is included (see Table IA.XXIII).

[53] The application process for EIDL was handled directly by the SBA. While agents could also assist PPP borrowers, the SBA indicated that they do not have any records regarding the use of agents in the PPP.

from zip code j is associated with a 1.1 ppt increase in the relative probability of an agent facilitating loans in both zip codes i and j.[54] This is another example of a specific pandemic relief decision being influenced by social connections.

Agents potentially had a strong financial incentive to convince borrowers to pursue pandemic aid, and their advice and assistance could have played an important role in encouraging or discouraging fraud. To assess how fraud varies across agents, we start by plotting a distribution of suspicious loan rates across agents, focusing on agents who facilitated at least 25 EIDL loans (712 agents). The blue line in Panel A of Figure 9 shows that suspicious loan rates are highly variable across agents. On the low end, 23.7% of agents have no suspicious loans, and on the other extreme, 10.5% and 4.8% of agents (75 and 34 agents, respectively) have at least 15% and 20% of their loans flagged as suspicious, respectively. For comparison, we also simulate what the distribution of fraud rates across agents would have been if the probability of a loan being flagged was independent of agents. The solid black line in the figure plots the distribution of simulated fraud rates across agents in the average of 100,000 simulations, assuming each loan has an independent probability of 5.1% of being flagged as suspicious, which is the average suspicious EIDL loan rate across all loans.[55] In the average simulation, only 1.4% (0.2%) of agents have at least 15% (20%) of their loans flagged as suspicious.

We next examine the role of agents in spreading fraud. For loans facilitated by agents with at least five previous loans (102,929 loans), we assess whether the suspicious loan rate among loans the agent previously facilitated is predictive of the likelihood that a subsequent loan that they facilitate is suspicious. A ten ppt increase in an agent's past suspicious EIDL loan rate is associated with a 5.08 ppt increase in the likelihood that the loan is flagged, which is a large effect compared to the unconditional flag rate of 5.46%. The left subpanel of Figure 9, Panel B graphically shows this relation as a binscatter. The right subpanel of Figure 9, Panel B shows that a similarly strong relationship exists even after zip code fixed effects are included.[56]

Overall, these results suggest that borrowers matched with agents at least in part based on social

[54] For this analysis, we estimate regressions in which the unit of observation is a zip code pair and restrict the analysis to agents that assist borrowers in multiple zip codes. Results are shown in Table IA.XXIV, Panel A.

[55] Figure IA.14 reruns the simulations assuming each loan has an independent probability equal to the fraud rate in its zip code of being flagged as suspicious. The results are very similar to those described below.

[56] Table IA.XXIV, Panel B shows equivalent results based on loan-level regressions using different thresholds on the number of previous loans required.

networks and that agents may have helped encourage or discourage EIDL fraud. This highlights another important channel through which social connections seemingly impact suspicious lending.

6 Potential Long-Term Effects

COVID relief fraud was widespread and potentially had long-lasting impacts on awareness, social norms, and information sharing related to fraud. In particular, one might wonder if COVID relief fraud was purely an isolated event or if the same people who engaged in COVID relief fraud could be actively seeking to engage in more recent post-COVID scams. To examine this question, we had research assistants manually scour the same large social media platform used for the analysis in Section 5.1 for groups encouraging potential suspicious financial activity in 2022 and 2023 using search terms such as "free money," "sauce," and "method." We find 127 suspicious public groups with a total of 79,418 unique members. The nature of the current scams is related to credit cards, banks, cryptocurrencies, wire transfers, gift cards, and electronic payment platforms (such as Cash App, PayPal, Square, Stripe, etc.).[57]

We examine whether social media users in pandemic relief groups with high PPP fraud rates are more likely to be in social media groups discussing suspicious financial activity in 2022–23 compared to users in pandemic relief groups with low fraud rates. Continued participation in suspicious groups need not be causally related to PPP fraud social media activity, but it indicates that the same individuals are continuing to pursue fraud opportunities using the same social media platform. Specifically, we analyze the 72 pandemic relief groups with at least 25 members matched to the PPP loan-level data from Section 5.1, split into quartiles based on the suspicious loan rates of their members. We then use user identifiers to see if members of the groups with high rates of PPP fraud are also more likely to be members of suspicious groups in 2022 and 2023. The left subpanel of Figure 10, Panel A shows the results. Users who were in the top quartile of pandemic relief groups by PPP fraud have a membership rate in 2022–23 suspicious groups of 9.6%, compared to 1.1% for the bottom quartile. This likely understates participation in suspicious post-COVID groups because we do not observe private groups, and our search for suspicious groups is not comprehensive.

[57] Examples of social media posts in 2022–23 discussing suspicious financial activity are provided in Exhibit IA.3. To focus on social media activity in 2022–23, groups are required to have started in 2022 or 2023 in order to be included in this analysis.

As a specific example of ongoing activity, we also explore groups that are focused on the Employee Retention Tax Credit (ERC) in 2022–23.[58] Citing a large influx of fraudulent applications due to "aggressive marketing to ineligible applicants," the IRS on September 14, 2023 paused the processing of new ERC claims until the end of 2023 (IRS, 2023; Wall Street Journal, 2023b). To understand whether members of suspicious pandemic relief social media groups are more likely to be discussing the ERC in 2022–23, we identify 32 public social media groups, with a total of 33,032 unique members, discussing the ERC in 2022–23.[59] We match the members of these groups to the members of pandemic relief social media groups as discussed above and examine whether social media users in groups with high PPP fraud rates are more likely to be in these groups. The right subpanel of Figure 10, Panel A presents the results. Users who were in the top quartile of pandemic relief groups by PPP fraud rates have a participation rate in ERC groups of 3.4%, compared to 0.5% for the bottom quartile.

To further understand connections between the ERC and pandemic relief fraud, we examine Google search activity across Designated Market Areas (DMA) based on data from Google Trends.[60] Figure 10, Panel B shows that the DMAs with high amounts of flagged PPP loans per capita have higher levels of search activity related to the ERC in 2022–23. The relation has a correlation of 0.587 with a t-statistic of 7.37.[61]

Overall, our findings suggest that social media remains an ongoing channel to promote fraudulent schemes. Since our analysis is based on only one social media platform, and other platforms may be less transparent, our exploratory look at potential current scams only scratches the surface of the interaction between social media and current suspicious activity.

7 Conclusion

In response to the COVID-19 pandemic, the U.S. government opened the relief floodgates with limited safeguards and growing evidence of fraud across multiple programs. This paper creates new

[58] The ERC was created early in the pandemic to incentivize businesses to keep employees on their payrolls. Starting in 2022 and accelerating in 2023, various consulting firms began heavily advertising that they could help clients claim the ERC in exchange for a commission of up to 25%. Through early March 2023, the IRS had paid $150 billion in ERC refunds, and total payments through July 2023 could be over $220 billion with another $120 billion in the pipeline (Wall Street Journal, 2023a).

[59] Examples of social media posts in 2022–23 discussing the ERC are provided in Exhibit IA.4.

[60] DMAs are the most granular geographic at which data is released by Google Trends and on average consist of 15 counties.

[61] See Figure IA.15 for additional analysis of data from Google Trends. Quarterly analysis of the connection between PPP fraud and search activity for the ERC shows that the relation strengthens beginning in the second quarter of 2022, which is when the amount of ERC refunds began increasing. The correlation between PPP fraud and search activity for the ERC in 2020–21 is much weaker. The relationship between PPP fraud and search activity related to the ERC in 2022–23 remains after controlling for past search activity related to the ERC. Additionally, search activity related to the ERC in 2022–23 is correlated with search activity in 2020–21 related to the PPP, Womply, and BlueAcorn.

measures for suspicious EIDL advances and excess unemployment insurance claims and finds strong geographic clustering in suspicious pandemic relief, both within the PPP and across programs despite variations in their target populations, application processes, and use of intermediaries. Social connections strongly predict geographic differences in suspicious lending rates and the spread of fraud over time.

The rapid spread of pandemic relief fraud is an important cautionary tale for government relief programs in an era of broad technological access and immediate information flow. Although there are strong arguments for why financial technology with reputational capital should deter and decrease fraud (Karpoff, 2021), in this context, financial technology and social connections seemingly accelerated a trickle of initial fraud in scattered geographic pockets in the early stages of the pandemic into a massive flow of broadening fraud across multiple government programs. Eliason et al. (2021) and Shi (2023) find that administrative actions and audits in health care are cost-effective for fraud prevention. The COVID-related programs illustrate what can happen when effective prevention policies are not applied in a timely manner. The rapid speed with which pandemic-related fraud grew to levels exceeding 50% in many zip codes is consistent with cascades in social norms (Bikhchandani, Hirshleifer, and Welch, 1998). Akerlof and Romer (1993) demonstrate that fraud can have large externalities, and Kedia and Philippon (2007) show how fraudulent accounting distorts the use of economic resources. In related research, Griffin, Kruger, and Mahajan (2024) find that the geographically concentrated pandemic fraud found in this paper distorted home prices, vehicle purchases, and consumer spending. This also highlights the growing role of rent-seeking (Zingales, 2015) at the intersection of government programs and the financial system. Perhaps most concerning, the rapid spread of pandemic fraud may have emboldened a social normalization (Sutherland, 1939) of the unabashed pursuit of “free money” from looting government programs. Against this backdrop, targeting more resources toward proactive and timely enforcement seems warranted.

References

Acemoglu, Daron, Vasco M. Carvalho, Asuman Ozdaglar, and Alireza Tahbaz-Salehi, 2012, The network origins of aggregate fluctuations, *Econometrica* 80, 1977–2016.

Akerlof, George A., and Paul M. Romer, 1993, Looting: The economic underworld of bankruptcy for profit, *Brookings Papers on Economic Activity* 1993, 1–73.

Allen, Linda, Lin Peng, and Yu Shan, 2024, Social networks as information conduits for online credit supply and demand, Working Paper.

Aman-Rana, Shan, Daniel W. Gingerich, and Sandip Sukhtankar, 2023, Screen now, save later? The trade-off between administrative ordeals and fraud, Working Paper.

Associated Press, 2023, The great grift: How billions in COVID-19 relief aid was stolen or wasted, Richard Lardner, Jennifer McDermott, and Aaron Kessler, June 11.

Autor, David, David Cho, Leland Crane, Mita Goldar, Byron Lutz, Joshua Montes, William Peterman, David Ratner, Daniel Villar, and Ahu Yildirmaz, 2022, The $800 billion Paycheck Protection Program: Where did the money go and why did it go there?, *Journal of Economic Perspectives* 36, 55–80.

Bailey, Michael, Rachel Cao, Theresa Kuchler, Johannes Stroebel, and Arlene Wong, 2018a, Social connectedness: Measurement, determinants, and effects, *Journal of Economic Perspectives* 32, 259–280.

Bailey, Michael, Ruiqing Cao, Theresa Kuchler, and Johannes Stroebel, 2018b, The economic effects of social networks: Evidence from the housing market, *Journal of Political Economy* 126, 2224–2276.

Bailey, Michael, Eduardo Dávila, Theresa Kuchler, and Johannes Stroebel, 2018c, House price beliefs and mortgage leverage choice, *The Review of Economic Studies* 86, 2403–2452.

Bailey, Michael, Patrick Farrell, Theresa Kuchler, and Johannes Stroebel, 2020, Social connectedness in urban areas, *Journal of Urban Economics* 118, 103264.

Bailey, Michael, Abhinav Gupta, Sebastian Hillenbrand, Theresa Kuchler, Robert Richmond, and Johannes Stroebel, 2021, International trade and social connectedness, *Journal of International Economics* 129, 103418.

Baker, Wayne E., and Robert R. Faulkner, 2003, Diffusion of fraud: Intermediate economic crime and investor dynamics, *Criminology* 41, 1173–1206.

Bartik, Alexander W, Zoe B Cullen, Edward L Glaeser, Michael Luca, Christopher T Stanton, and Adi Sunderam, 2023, When should public programs be privately administered? Theory and evidence from the Paycheck Protection Program, Working Paper.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch, 1998, Learning from the behavior of others: Conformity, fads, and informational cascades, *Journal of Economic Perspectives* 12, 151–170.

Chernenko, Sergey, and David S. Scharfstein, 2024, Racial disparities in the Paycheck Protection Program, *Journal of Financial Economics* Forthcoming.

Chetty, Raj, John N. Friedman, and Emmanuel Saez, 2013, Using differences in knowledge across neighborhoods to uncover the impacts of the EITC on earnings, *The American Economic Review* 103, 2683–2721.

Chetty, Raj, John N Friedman, Michael Stepner, and Opportunity Insights Team, 2023, The economic impacts of COVID-19: Evidence from a new public database built using private sector data, *The Quarterly Journal of Economics* 139, 829–889.

Chetty, Raj, Matthew O. Jackson, Theresa Kuchler, Johannes Stroebel, Nathaniel Hendren, Robert B. Fluegge, Sara Gong, Federico Gonzalez, Armelle Grondin, Matthew Jacob, Drew Johnston, Martin Koenen, Eduardo Laguna-Muggenburg, Florian Mudekereza, Tom Rutter, Nicolaj Thor, Wilbur Townsend, Ruby Zhang, Mike Bailey, Pablo Barberá, Monica Bhole, and Nils Wernerfelt, 2022a, Social capital I: measurement and associations with economic mobility, *Nature* 608, 108–121.

Chetty, Raj, Matthew O. Jackson, Theresa Kuchler, Johannes Stroebel, Nathaniel Hendren, Robert B. Fluegge, Sara Gong, Federico Gonzalez, Armelle Grondin, Matthew Jacob, Drew Johnston, Martin Koenen, Eduardo Laguna-Muggenburg, Florian Mudekereza, Tom Rutter, Nicolaj Thor, Wilbur Townsend, Ruby Zhang, Mike Bailey, Pablo Barberá, Monica Bhole, and Nils Wernerfelt, 2022b, Social capital II: determinants of economic connectedness, *Nature* 608, 122–134.

Cinelli, Matteo, Gianmarco De Francisci Morales, Alessandro Galeazzi, Walter Quattrociocchi, and Michele Starnini, 2021, The echo chamber effect on social media, *Proceedings of the National Academy of Sciences* 118, e2023301118.

Cramer, Kim Fe, and Naz Koont, 2021, Peer effects in deposit markets, Working Paper.

de Soyres, Francois, Ana Maria Santacreu, and Henry L. Young, 2023, Demand-supply imbalance during the COVID-19 pandemic: The role of fiscal policy, *Federal Reserve Bank of St. Louis Review* First Quarter 2023, 21–50.

Deason, Stephen, Shivaram Rajgopal, Gregory Waymire, and Roger White, 2015, Who gets swindled in Ponzi schemes?, Working Paper.

Del Vicario, Michela, Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, Guido Caldarelli, H. Eugene Stanley, and Walter Quattrociocchi, 2016, The spreading of misinformation online, *Proceedings of the National Academy of Sciences* 113, 554–559.

Dimmock, Stephen G., William C. Gerken, and Nathaniel P. Graham, 2018, Is fraud contagious? Coworker influence on misconduct by financial advisors, *The Journal of Finance* 73, 1417–1450.

Duflo, Esther, and Emmanuel Saez, 2003, The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment, *The Quarterly Journal of Economics* 118, 815–842.

Décary-Hêtu, and Carlo Morselli, 2011, Gang presence in social network sites, *International Journal of Cyber Criminology* 5, 876.

Eliason, Paul J, Riley J League, Jetson Leder-Luis, Ryan C McDevitt, and James W Roberts, 2021, Ambulance taxis: The impact of regulation and litigation on health care fraud, Working Paper 29491, National Bureau of Economic Research.

Erel, Isil, and Jack Liebersohn, 2022, Can FinTech reduce disparities in access to finance? Evidence from the Paycheck Protection Program, *Journal of Financial Economics* 146, 90–118.

Fang, Lei, and Mark S. Seasholes, 2004, Correlated trading and location, *The Journal of Finance* 59, 2117–2144.

Faulkender, Michael W., Robert Jackman, and Stephen Miran, 2021, The job preservation effects of Paycheck Protection Program loans, Working Paper.

Giroud, Xavier, and Holger M. Mueller, 2019, Firms' internal networks and local economic shocks, *American Economic Review* 109, 3617–49.

Glaeser, Edward L., Bruce Sacerdote, and José A. Scheinkman, 1996, Crime and social interactions, *The Quarterly Journal of Economics* 111, 507–548.

Granja, João, Christos Makridis, Constantine Yannelis, and Eric Zwick, 2022, Did the Paycheck Protection Program hit the target?, *Journal of Financial Economics* 145, 725–761.

Griffin, John M., Samuel Kruger, and Prateek Mahajan, 2023, Did FinTech lenders facilitate PPP fraud?, *The Journal of Finance* 78, 1777–1827.

Griffin, John M., Samuel Kruger, and Prateek Mahajan, 2024, Did pandemic relief fraud inflate house prices?, Working Paper.

Griffin, John M., Samuel Kruger, and Gonzalo Maturana, 2019, Personal infidelity and professional conduct in 4 settings, *Proceedings of the National Academy of Sciences* 116, 16268–16273.

Grullon, Gustavo, George Kanatas, and James Weston, 2010, Religion and corporate (mis)behavior, Working Paper.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales, 2013, The determinants of attitudes toward strategic default on mortgages, *The Journal of Finance* 68, 1473–1515.

Gurun, Umit G, Noah Stoffman, and Scott E Yonker, 2017, Trust busting: The effect of fraud on investor behavior, *The Review of Financial Studies* 31, 1341–1376.

Hirshleifer, David, 2020, Presidential address: Social transmission bias in economics and finance, *The Journal of Finance* 75, 1779–1831.

Holzman, Eric R., Brian P. Miller, and Brian M. Williams, 2021, The local spillover effect of corporate accounting misconduct: Evidence from city crime rates, *Contemporary Accounting Research* 38, 1542–1580.

Hong, Harrison, Jeffrey D. Kubik, and Jeremy C. Stein, 2004, Social interaction and stock-market participation, *The Journal of Finance* 59, 137–163.

Howell, Sabrina T., Theresa Kuchler, David Snitkof, Johannes Stroebe, and Jun Wong, 2024, Lender automation and racial disparities in credit access, *The Journal of Finance* 79, 1457–1512.

Hu, Zhongchen, 2021, Social interactions and households' flood insurance decisions, *Journal of Financial Economics* 144, 414–432.

Humanitarian Data Exchange, 2021, Facebook social connectedness index, data notes.

IRS, 2023, To protect taxpayers from scams, IRS orders immediate stop to new employee retention credit processing amid surge of questionable claims; concerns from tax pros, September 14.

Jordà, Òscar, and Fernanda Nechio, 2023, Inflation and wage growth since the pandemic, *European Economic Review* 156, 104474.

Karpoff, Jonathan M., 2021, The future of financial fraud, *Journal of Corporate Finance* 66, 101694.

Kedia, Simi, and Thomas Philippon, 2007, The Economics of Fraudulent Accounting, *The Review of Financial Studies* 22, 2169–2199.

Khetan, Umang, Jetson Leder-Luis, Jialan Wang, and Yunrong Zhou, 2024, Unemployment insurance fraud in the debit card market, Working Paper 32527, National Bureau of Economic Research.

Kuchler, Theresa, Yan Li, Lin Peng, Johannes Stroebel, and Dexin Zhou, 2022, Social proximity to capital: Implications for investors and firms, *The Review of Financial Studies* 35, 2743–2789.

Kuchler, Theresa, Dominic Russel, and Johannes Stroebel, 2022, The geographic spread of COVID-19 correlates with the structure of social networks as measured by facebook, *Journal of Urban Economics* 127, 103314.

Kuchler, Theresa, and Johannes Stroebel, 2021, Social finance, *Annual Review of Financial Economics* 13, 37–55.

Levi, Michael, 2008, Organized fraud and organizing frauds: Unpacking research on networks and organization, *Criminology & Criminal Justice* 8, 389–419.

Levy, Ro'ee, 2021, Social media, news consumption, and polarization: Evidence from a field experiment, *American Economic Review* 111, 831–70.

Manski, Charles F., 1993, Identification of endogenous social effects: The reflection problem, *The Review of Economic Studies* 60, 531–542.

Matsueda, Ross L, 1988, The current state of differential association theory, *Crime & Delinquency* 34, 277–306.

Maturana, Gonzalo, and Jordan Nickerson, 2018, Teachers Teaching Teachers: The Role of Workplace Peer Effects in Financial Decisions, *The Review of Financial Studies* 32, 3920–3957.

McGloin, Jean Marie, and Kyle J. Thomas, 2019, Peer influence and delinquency, *Annual Review of Criminology* 2, 241–264.

McGloin, Jean Marie, Kyle J. Thomas, and Christopher J. Sullivan, 2019, Quantifying the likelihood of false positives: Using sensitivity analysis to bound statistical inference, *Journal of Quantitative Criminology* 35, 631–662.

Müller, Karsten, and Carlo Schwarz, 2020, Fanning the flames of hate: Social media and hate crime, *Journal of the European Economic Association* 19, 2131–2167.

Nash, Rebecca, Martin Bouchard, and Aili Malm, 2013, Investing in people: The role of social networks in the diffusion of a large-scale fraud, *Social Networks* 35, 686–698.

NBC News, 2022, 'Biggest fraud in a generation': The looting of the Covid relief plan known as ppp, Ken Dilanian and Laura Strickler, March 28.

O'Malley, A. James, Thomas A Bubolz, and Jonathan S Skinner, 2023, The diffusion of health care fraud: A bipartite network analysis, *Social Science & Medicine* 327, 115927.

Oster, Emily, 2019, Unobservable selection and coefficient stability: Theory and evidence, *Journal of Business & Economic Statistics* 37, 187–204.

Parsons, Christopher A., Johan Sulaeman, and Sheridan Titman, 2018, The geography of financial misconduct, *Journal of Finance* 73, 2087–2137.

ProPublica, 2021, How unemployment insurance fraud exploded during the pandemic, Cezary Podkul, July 26.

Puri, Neha, Eric A. Coomes, Hourmazd Haghbayan, and Keith and Gunaratne, 2020, Social media and vaccine hesitancy: New updates for the era of COVID-19 and globalized infectious diseases, *Human Vaccines & Immunotherapeutics* 16, 2586–2593.

Shekhar, Shubhranshu, Jetson Leder-Luis, and Leman Akoglu, 2023, Unsupervised machine learning for explainable health care fraud detection, Working Paper 30946, National Bureau of Economic Research.

Shi, Maggie, 2023, Monitoring for Waste: Evidence from Medicare Audits*, *The Quarterly Journal of Economics* 139, 993–1049.

Silverman, Elaine, and Jonathan Skinner, 2004, Medicare upcoding and hospital ownership, *Journal of Health Economics* 23, 369–389.

Suarez-Lledo, Victor, and Javier Alvarez-Galvez, 2021, Prevalence of health misinformation on social media: Systematic review, *Journal of Medical Internet Research* 23, e17187.

Sutherland, Edwin Hardin, 1939, *Principles of criminology* (J.B. Lippincott Company).

Van Vlasselaer, Véronique, Tina Eliassi-Rad, Leman Akoglu, Monique Snoeck, and Bart Baesens, 2017, Gotcha! Network-based fraud detection for social security fraud, *Management Science* 63, 3090–3110.

Vosoughi, Soroush, Deb Roy, and Sinan Aral, 2018, The spread of true and false news online, *Science* 359, 1146–1151.

Wall Street Journal, 2023a, Inside a sales army turning a tax break into a modern-day gold rush, Ruth Simon and Richard Rubin, September 4.

Wall Street Journal, 2023b, IRS shuts door on new pandemic tax credit claims until at least 2024, Richard Rubin, September 14.

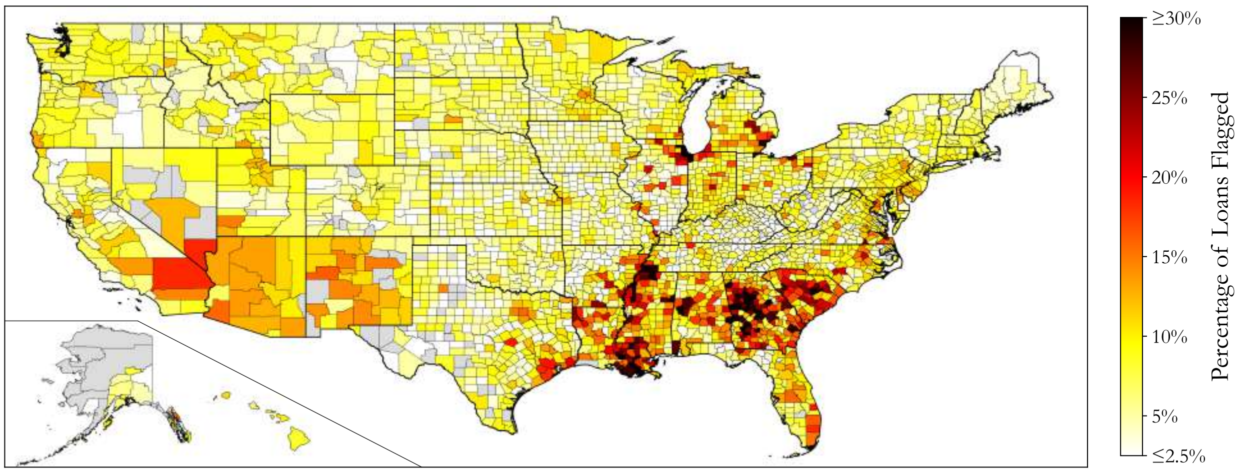
Wilson, Riley, 2022, The impact of social networks on eitc claiming behavior, *The Review of Economics and Statistics* 104, 929–945.

Zingales, Luigi, 2015, Presidential address: Does finance benefit society?, *Journal of Finance* 70, 1327–1363.

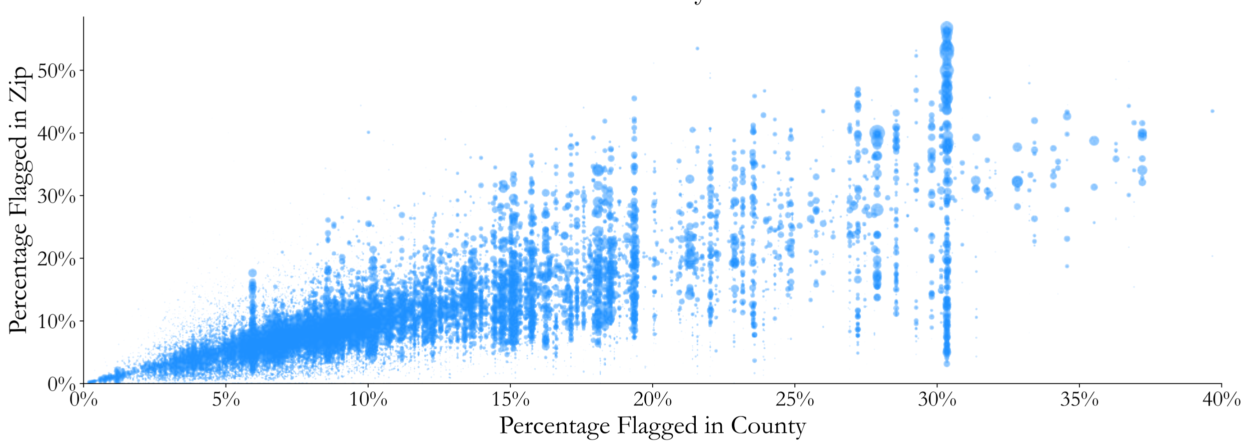
Figure 1. Geography of Flagged Loans

This figure shows geographic variation in the percentage of flagged loans. Panel A shows the percentage of flagged loans in each county, Panel B shows within county variation, and Panel C shows the distributions of flag rates across counties and over time for loans originated by FinTech and traditional lenders. In Panel A, counties are colored based on the color bar to the right of the map, and counties with fewer than 100 loans are colored grey. In Panel B, the percentage of flagged loans in each zip code is shown on the vertical axis and the percentage of flagged loans in the corresponding county on the horizontal axis. Dots are sized based on the number of loans in the zip code. Zip codes with at least 25 loans are shown. In Panel C, the boxplots are weighted by the number of loans in the given county-lender type-month cell. County-lender type-month cells with fewer than 25 loans are excluded. The flag rates in five specific counties are also shown.

Panel A. County Level Heatmap



Panel B. Within County Variation



Panel C. By Month and Lender Type

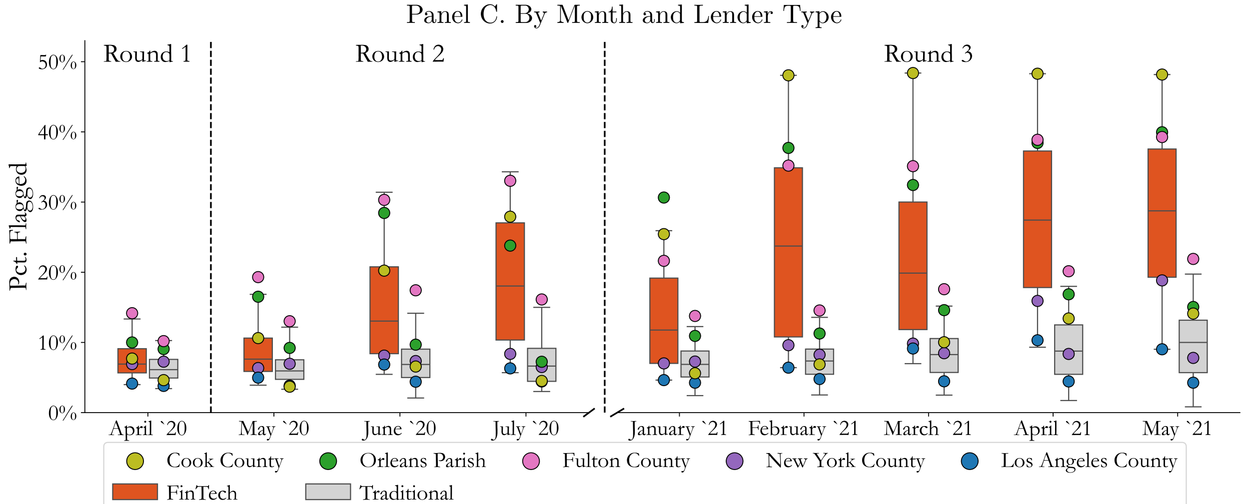
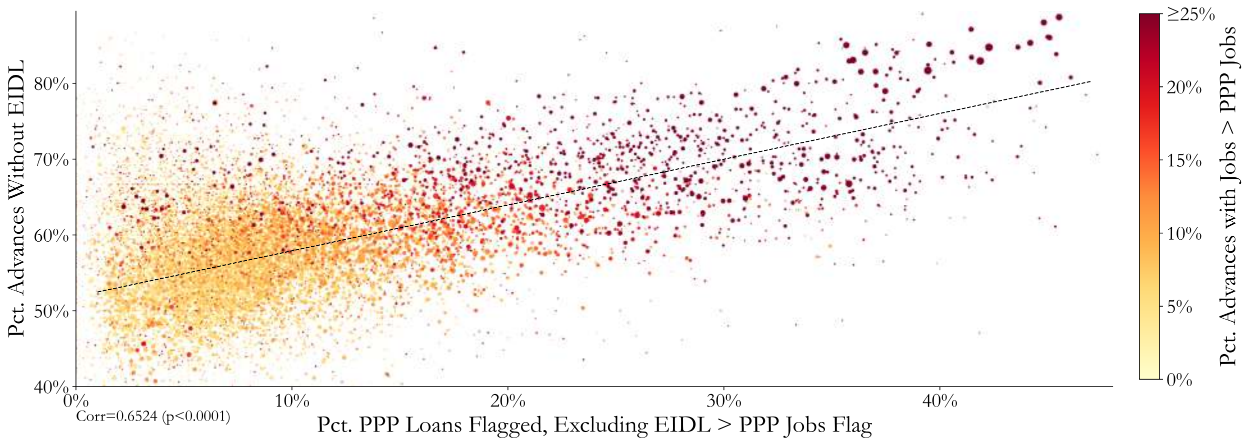


Figure 2. Correlation Between Pandemic Relief Programs

This figure shows the geographic correlations between suspicious activity across programs. Panel A shows the relationship between the percentage of EIDL Advances without a corresponding EIDL and the percentage of PPP loans that are flagged. Each dot is a zip code, is colored based on the percentage of PPP loans that are flagged by the EIDL > PPP Jobs flag, and is sized based on the number of EIDL Advances. The dashed line is a linear fit and the correlation is shown in the bottom left corner. The top subpanel of Panel B shows the effect of a one standard deviation increase in PPP fraud rates on initial unemployment insurance claims based on monthly regression that controls for overall PPP loans per capita, state fixed effect, and state fixed effects interacted with demographics. The regressions are weighted by the number of individuals in the labor force as of December 2019. The demographic control variables are the percentage of adults with at least a bachelor's degree, median income, pre-pandemic unemployment rate, population density, percentage non-white, and poverty rate. Standard errors are clustered by state. The two smaller subpanels show the effects of PPP fraud on both the level and monthly change in employment based on data from the Economic Tracker by Opportunity Insights (described in Chetty et al. (2023)).

Panel A. Correlation Between EIDL Advance and PPP



Panel B. Correlation Between Unemployment Insurance and PPP

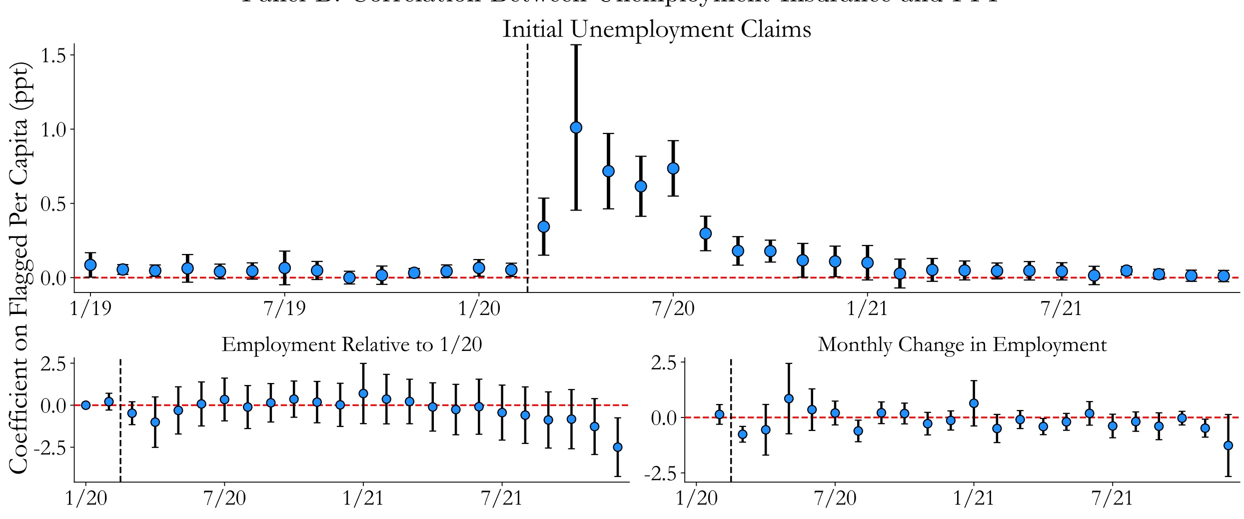


Figure 3. Network of Fraud

This figure shows the network of social connections between three high-fraud CSAs: Chicago, Atlanta, and New Orleans. Chicago zip codes are shown on the left side, Atlanta on the right side, and New Orleans on the bottom. Each node is a zip code and the edges are connections between pairs of zip codes. Within-CSA connections are excluded. The color of the node is based on the percentage of PPP loans that are flagged. The width of each edge represents the strength of social connections between the zip codes based on the Social Connectedness Index created by Bailey et al. (2018a, 2020). Zip codes are split into ten groups within each CSA based on their level of fraud. The red arrows point towards the higher fraud nodes in each CSA.

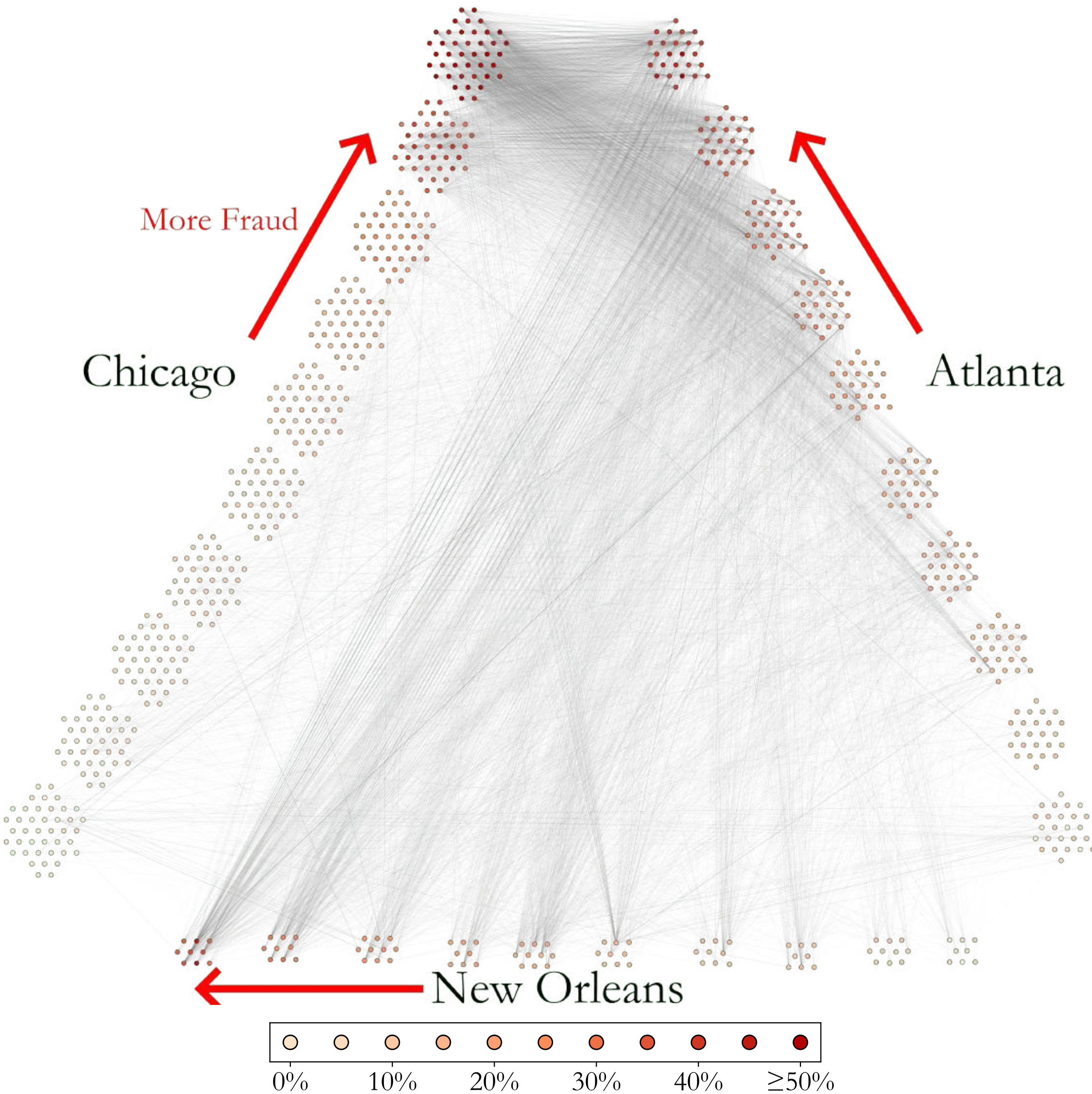
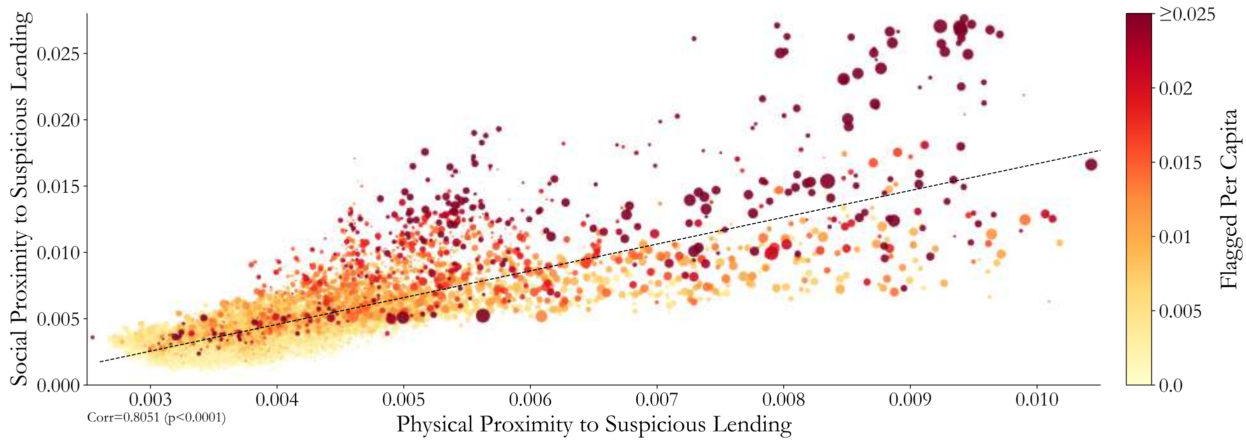


Figure 4. Social and Physical Proximity to Suspicious Lending

This figure shows the relationship between social/physical proximity to suspicious lending and flagged per capita across zip codes. In Panel A, each dot represents a zip code, is sized based on the number of loans in the zip code, and is colored based on the ratio of the number of flagged loans to the population of the zip code (based on the color bar). Panels B show binscatters based on data at the zip code level. Both subpanels of Panel B include county fixed effects and controls for log population density, percentage non-white, the log of average household income, poverty rate, pre-pandemic unemployment, percentage with college education, the share of Facebook friends within 50 and 150 miles, and the FinTech market share of PPP loans in the zip code. The left (right) subpanel of Panel B additionally control for physical (social) proximity to suspicious lending and show the relationship between flagged loans per capita and social (physical) proximity to suspicious lending. To have a nationally representative estimate, all panels use weighted least squares (WLS) regressions with the weight being the number of PPP loans in each zip code. Zip codes are filtered to those with at least 25 loans and the dashed lines are linear fits.

Panel A. Social Proximity vs. Physical Proximity to Suspicious Lending



Panel B. Binscatter with Controls and County Fixed Effects

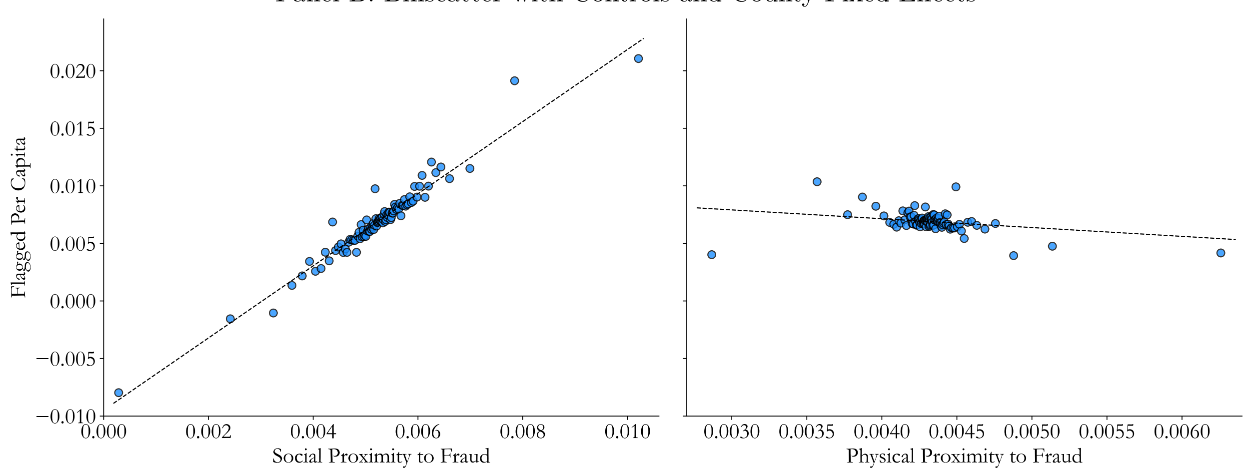


Figure 5. Heterogeneity in Effect of Social and Physical Proximity

This figure shows heterogeneity in the relationship between social/physical proximity to suspicious lending and flagged per capita across demographic splits. Panel A shows heterogeneity in the effect of social proximity and Panel B in physical proximity. The regressions include county fixed effects and control for log population density, percentage non-white, the log of average household income, poverty rate, pre-pandemic unemployment, percentage with college education, the share of Facebook friends within 50 and 150 miles, and the FinTech market share of PPP loans in the zip code. The splits are based on the median value of the demographic across all zip codes. The error bars represent 95% confidence intervals based on robust standard errors that are clustered by county. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the number of PPP loans in each zip code. Zip codes are filtered to those with at least 25 loans.

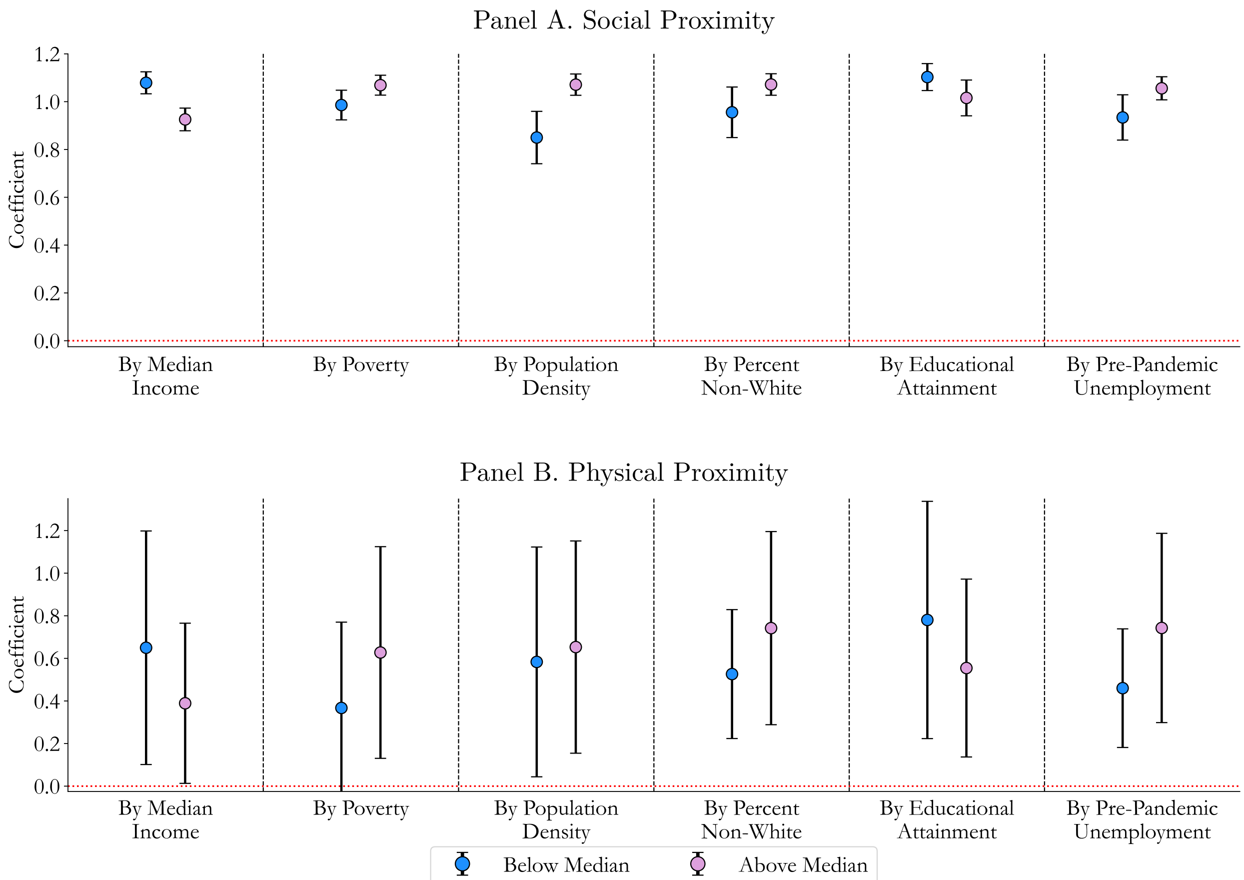
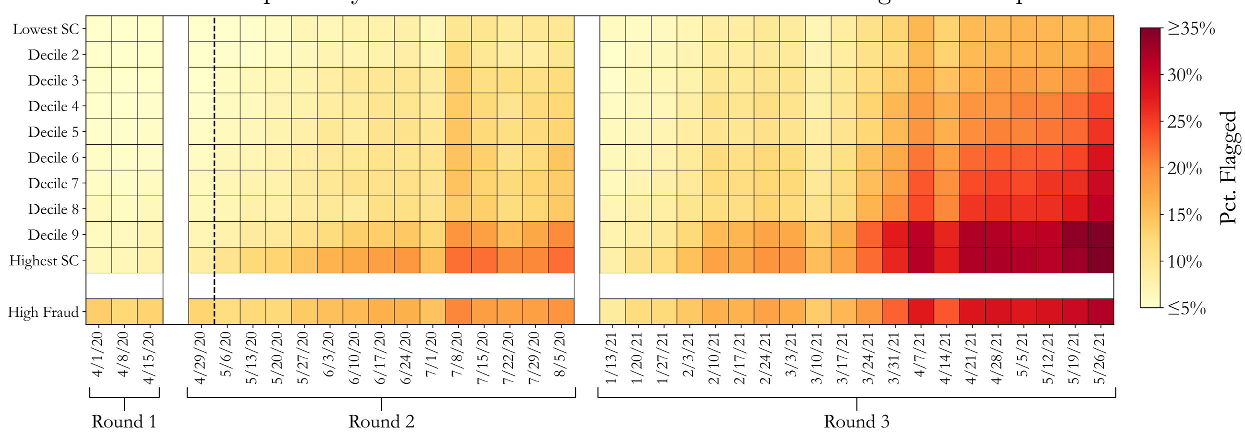


Figure 6. Spread of Fraud Over Time

This figure shows the spread of suspicious lending over time. Zip codes are sorted based on their flag rates during the first month of the program and the top zip codes that collectively represent ten percent of lending during the first month are the initial high fraud zip codes. The remaining zip codes are split into deciles (Panel A) or terciles (Panel B) based on their social connectedness to the initial high fraud zip codes. Specifically, for each zip code i, $Social\ Connectedness_{i,Initial\ High\ Fraud} = \sum_{j \in Initial\ High\ Fraud} Population_j \times SocialConnectedness_{i,j}$. Panel A then shows the flag rate during each week in the initial high fraud zip codes and each decile of zip codes. The color bar to the right of the figure provides the interpretation of the coloring scheme. In Panel B, each zip code in the top tercile of social connectedness to initial high fraud zip codes is matched (with replacement) to a zip code in the bottom tercile that has similar flag rates during the first month of the program. The flag rates during each week in the top tercile of zip codes by social connectedness to initial high fraud zip codes and the matched zip codes in the bottom tercile are shown. The dotted lines in Panel B represent 95% confidence intervals based on standard errors clustered at the zip code level. Zip codes with fewer than ten loans in the first month are dropped. Each decile or tercile is formed such that it has an equal number of loans in the first month. The p-value for the difference in percentage flagged during the first month of the program in zip codes in the top tercile of social connectedness and the matched zip codes in the bottom tercile is noted in Panel B.

Panel A. Spread by Decile of Social Connectedness to Initial High Fraud Zip Codes



Panel B. Matched on Initial Fraud Rates

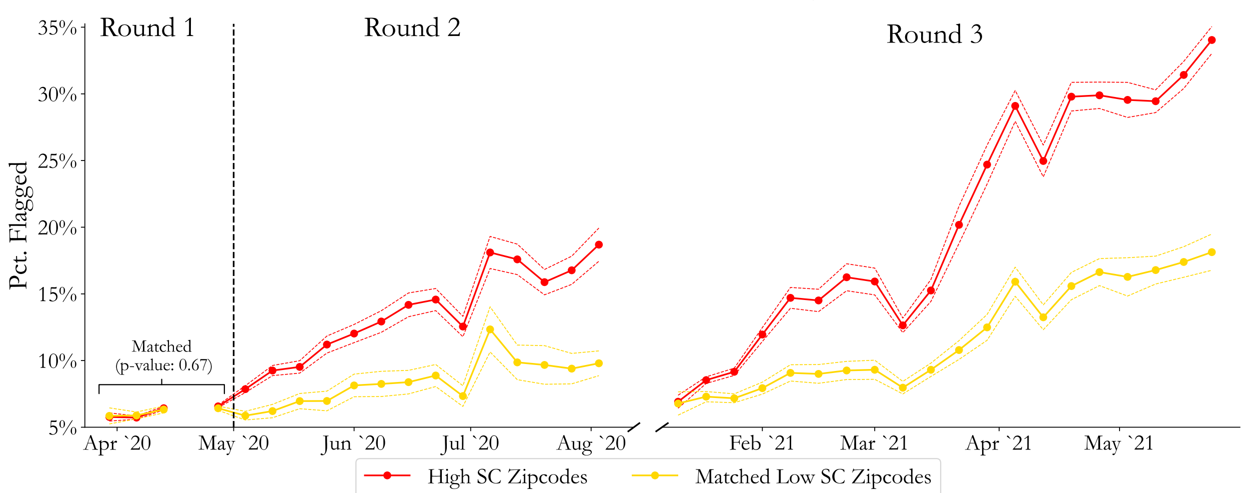
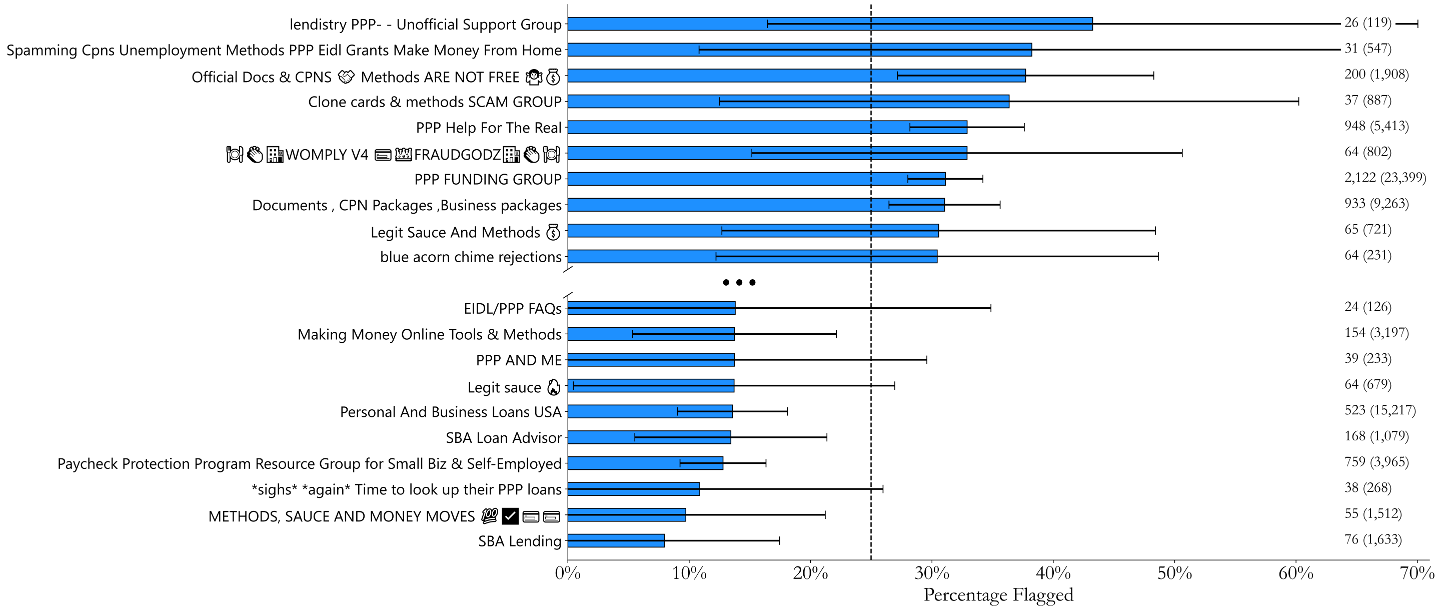


Figure 7. Social Media Groups

This figure examines social media groups on a large social media platform with discussions regarding the PPP and any other pandemic programs. Panel A shows the percentage of PPP loans to members of each group that are flagged. Groups with less than 25 loans matched to their members are excluded. For ease of viewing, we only show the 10 groups with the highest and lowest flag rates among the loans matched to their members; Figure IA.11 shows a version of Panel A with all 72 groups. The number of members of each group that are matched to a PPP loan and, in parentheses, the number of members in each group are shown to the right of each bar. The error bars represent 95% confidence intervals with Bonferroni correction for multiple testing. The dashed line is the average flag rate across loans matched to members of any of the groups. Panel B shows the 200 most common words being used in the bottom (left) and top (right) ten groups by the percentage of their members' PPP loans that are flagged. Words that are common in general, also known as stop words, (e.g., "the," "he," "she," and "it") are excluded. Additionally, names of programs and other common words in the pandemic relief context (e.g., "PPP," "EIDL," "unemployment," "borrower," and "lender") are excluded. The size of the words is determined by the frequency of their usage in each set of groups. The words shown in red are mainly used for nefarious purposes. For five of the larger groups, the full universe of posts is not available, and we are only able to collect around 1,700 posts per group.

Panel A. Percentage Flagged in Each Group



Panel B. Wordcloud of Bottom (Left) and Top (Right) Ten Groups

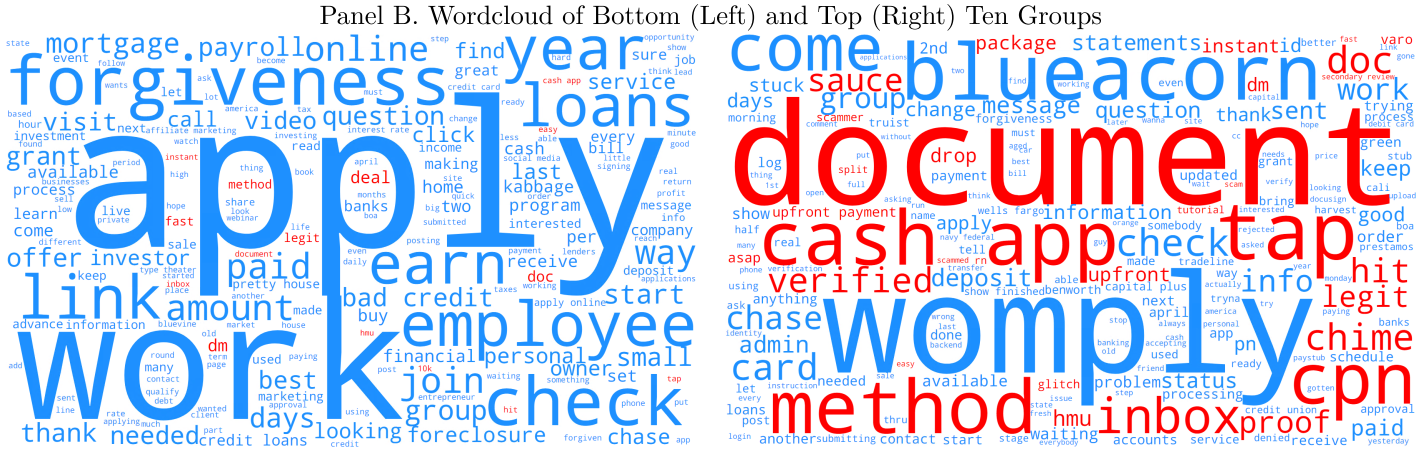


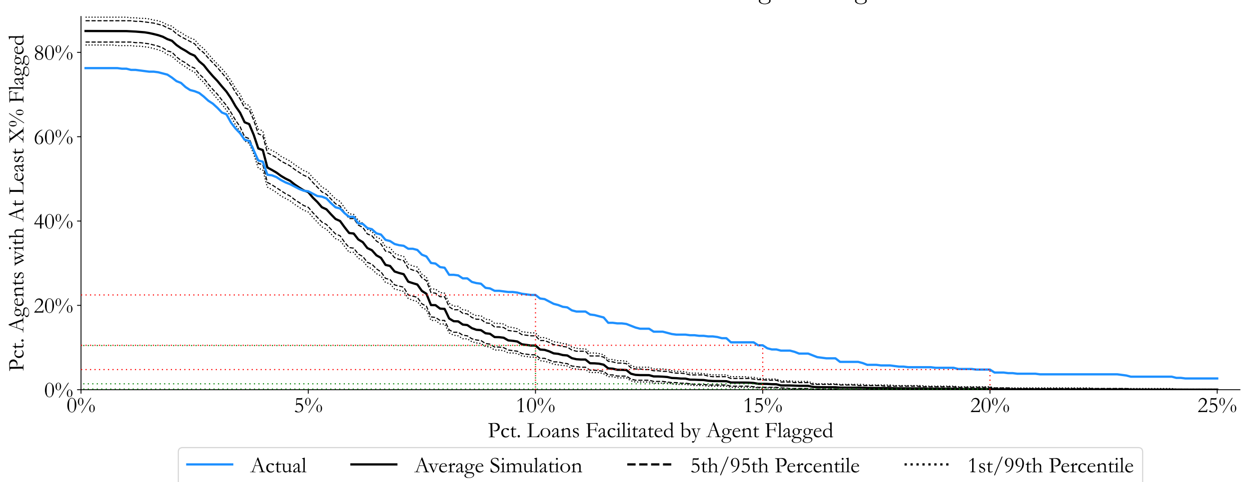
Figure 8. Clustering Along Loan Features

This figure examines clustering along loan features and whether clusters tend to be located in socially connected counties. Panel A shows two specific examples of loan clusters based on Griffin, Kruger, and Mahajan (2023), and Panel B shows the average effect across all clusters identified for Cross River, Womply, and BlueAcorn. Each cluster is defined by a combination of three loan features: the industry of the business receiving the loan, the loan amount, and the number of jobs reported. Both panels show the relationship between the percentage of a given lender's loans in each county that match along all three loan features and the social connectedness of the county to the central county of each cluster. Binscatters are shown. Each binscatter is run seperately, is weighted by the number of total loans by the lender in each county, and includes a state fixed effect. In Panel A, two clusters of loans originated by Cross River are consider. For both of these clusters, Cook County (Chicago) is the central county since it received the majority of loans matching these clusters. For Panel B, we determine clusters of loans by identifying combinations of industry, loan amount, and job reported with at least 100 loans originated by a single lender in a single county, and designate the county where the lender has the most loans matching each combination of loan features as the central county for each cluster's network. We define industry based on 4-digit NAICS codes and round loan amounts to the nearest $100 to allow for small differences. We exclude loans where the loan amount is at the maximum allowed per employee to avoid loans that may cluster at this point mechanically. Binscatters for each cluster are generated as in Panel A, and the average of the binscatters across all clusters identified for each lender is shown. In both panels, the blue dots/lines are based on loans by the focal lender, orange based on loans by other FinTech lenders, and green based on loans by traditional lenders.

Figure 9. EIDL Agents

This figure shows the complementary cumulative distribution function of agent flag rates in the EIDL program (Panel A) and the relation between an agent's past flag rate and the likelihood that their subsequent loan is flagged (Panel B). Loans are considered flagged if the loan is flagged by at least one of the business registry, multiple loans, or EIDL > PPP jobs flags (see Griffin, Kruger, and Mahajan (2023) for details on these flags). In Panel A, the percentage of agents with at least X% of their loans flagged in the actual data is shown by the blue curve, and the average percentage of agents with at least X% of their loans flagged in the 100,000 simulations is shown as the black curve. The region enclosed by dashed (dotted) black lines represents the percentage of agents with at least X% of their loans flagged in the middle 95% (99%) of simulations. The simulations are generated under the assumption that each loan has an independent probability of being flagged equal to the national flag rate. X% takes values between 0.1% and 25% in 0.1 ppt increments. Only agents with at least 25 loans are considered. In Panel B, binscatters are shown. The left subpanel shows the overall relationship and the right subpanel shows the relationship within zip codes by including zip code fixed effects. The dashed lines in Panel B are linear fits. Agents with at least five previous loans are considered.

Panel A. Cumulative Distribution of Agent Flag Rates



Panel B. Effect of Agent's Past Flag Rate on Likelihood Loan is Flagged

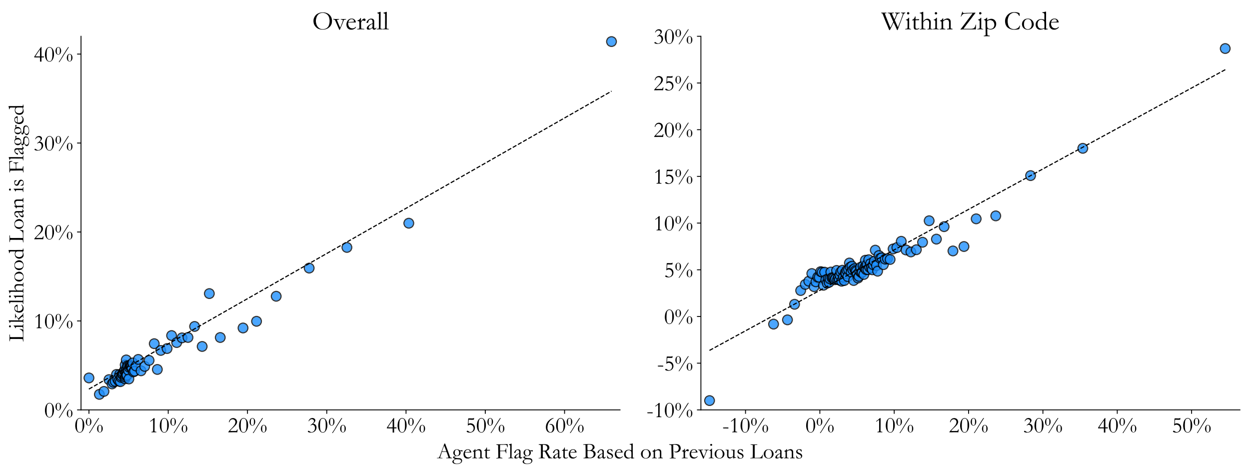
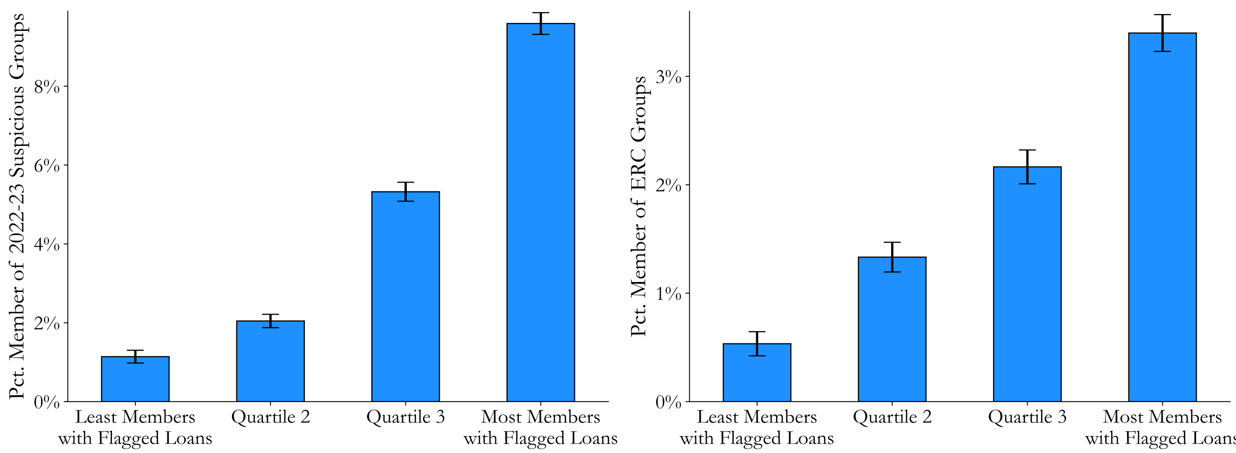


Figure 10. Connections to Activity in 2022–2023

This figure examines connections between suspicious activity during the pandemic and activity in 2022-23. Panel A examines if members of suspicious pandemic relief social media groups are more likely to also be members of 2022-23 social media groups with discussions related to potentially suspicious financial activity (left subpanel) or related to the Employee Retention Tax Credit [ERC] (right subpanel). The 72 pandemic relief social media groups from Figure 7 are split into quartiles based on the percentage of their members' PPP loans that are flagged. The left subpanel is based on 127 groups and the right subpanel is based on 32 groups, all of which are on the same platform as the groups in Figure 7. The bars show the percentage of pandemic relief social media group members who are also members of the 2022-23 social media groups. The error bars are 95% confidence intervals. Panel B examines the connection between PPP fraud and Google searches related to the ERC in 2022-23. Each dot is a Designated Market Area (DMA), the dashed line is a linear fit, and the correlation is shown in the bottom left corner.

Panel A. 2022-23 Social Media Groups Related to Suspicious Financial Activity and ERC



Panel B. Google Trends, Interest in ERC

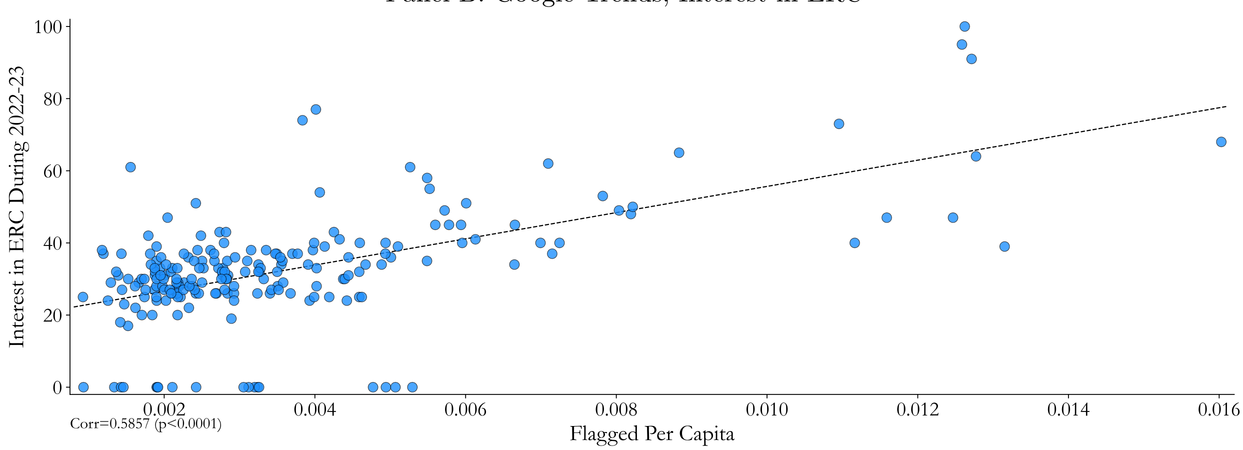


Table 1. Summary Statistics

This table provides summary statistics. Panel A provides summary statistics for variables at the zip code level and Panel B at the county level. *Pct. PPP Flagged* is the percentage of PPP loans that are flagged by at least one of the four primary flags developed by Griffin, Kruger, and Mahajan (2023). *PPP Loans Per Capita* is the ratio of PPP loans to the population of the zip code/county. *Flagged PPP Per Capita* is the ratio of PPP loans flagged by at least one of the four primary flags developed by Griffin, Kruger, and Mahajan (2023) to the population of the zip code/county. *Social Proximity to Suspicious PPP Lending* is the weighted average *Flagged PPP Per Capita* in each zip code/county's socially proximate zip codes/counties, with the weight being the Social Connectedness Index (from Bailey et al. (2018a, 2020)) between pairs of zip codes/counties. *Physical Proximity to Suspicious PPP Lending* is the weighted average *Flagged PPP Per Capita* in each zip code/county's physically proximate zip codes/counties, with the weight being the inverse distance between pairs of zip codes/counties. *Pct. EIDL Advance Without Loan* is the percentage of EIDL Advances in the zip code that do not have a corresponding EIDL loan. *Pct. EIDL Advance with Jobs > PPP Jobs* is the percentage of EIDL Advances recipients with more than three additional jobs implied by their grant amount than they reported on their PPP application. *Excess Share* is the share of first-draw business loans that exceed the Census establishment counts for a given industry in a particular county. *Excess Share Per Capita* is the ratio of the number of first-draw business loans that exceed the Census establishment counts for a given industry in a particular county to the population of the county. *UI Claims Divided by Labor Force* is the ratio of the number of initial UI claims filed between March and December 2020 divided by the size of the county's labor force. The statistics are weighted by number of PPP loans for *Pct. PPP Flagged* and *Excess Share*; by population for *PPP Loans Per Capita*, *Flagged PPP Per Capita*, *Social Proximity to Suspicious PPP Lending*, *Physical Proximity to Suspicious PPP Lending*, and *Excess Share Per Capita*; by number of EIDL Advances for *Pct. EIDL Advance Without Loan* and *Pct. EIDL Advance with Jobs > PPP Jobs*; and by individuals in the labor force for *UI Claims Divided by Labor Force.*

Panel A. Zip Code Level

	Mean	SD	p25	p50	p75	N
Pct. PPP Flagged	0.1240	0.0927	0.0663	0.0946	0.1489	24,325
PPP Loans Per Capita	0.0354	0.0270	0.0217	0.0296	0.0416	24,325
Flagged PPP Per Capita	0.0044	0.0060	0.0017	0.0028	0.0047	24,325
Social Proximity to Suspicious PPP Lending	0.0044	0.0026	0.0030	0.0037	0.0049	20,496
Physical Proximity to Suspicious PPP Lending	0.0040	0.0011	0.0034	0.0038	0.0043	20,496
Pct. EIDL Advances Without Loan	0.5867	0.0778	0.5368	0.5800	0.6270	17,040
Pct. EIDL Advances with Jobs > PPP Jobs	0.1143	0.1231	0.0500	0.0741	0.1223	17,040

Panel B. County Level

	Mean	SD	p25	p50	p75	N
Pct. PPP Flagged	0.1230	0.0729	0.0702	0.0983	0.1573	3,210
PPP Loans Per Capita	0.0349	0.0158	0.0249	0.0322	0.0413	3,210
Flagged PPP Per Capita	0.0043	0.0041	0.0020	0.0029	0.0048	3,210
Excess Share	0.2194	0.1341	0.1143	0.1846	0.3062	3,171
Excess Share Per Capita	0.0041	0.0043	0.0016	0.0027	0.0048	3,171
Social Proximity to Suspicious PPP Lending	0.0033	0.0012	0.0027	0.0031	0.0036	3,198
Physical Proximity to Suspicious PPP Lending	0.0032	0.0004	0.0029	0.0032	0.0034	3,198
UI Claims Divided by Labor Force	0.4244	0.2260	0.2590	0.3553	0.5093	2,199

Table 2. Social Proximity to Suspicious Lending

This table examines the zip code-level relationship between social proximity to fraud and the number of flagged loans per capita. The controls are the log of population density, percentage non-white, log of average household income, poverty rate, pre-pandemic unemployment, percentage with college education, the shares of the friends of Facebook users in the zip code who live within 50 and 150 miles of them, and the FinTech market share of PPP loans in the zip code. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the number of PPP loans in each zip code. All variables (both independent and dependent) are standardized at the zip code level to have a mean of 0 and a standard deviation of 1. Zip codes are filtered to those with at least 25 loans. Fixed effects and controls are as indicated at the bottom of each column. Robust standard errors are clustered by county.

Panel A. Social and Physical Proximity

Dep. Variable: Flagged By At Least One Primary Measure Per Capita in Zip Code

	(1)	(2)	(3)	(4)
Social Proximity to	1.110***	1.071***		1.116***
Suspicious Lending	(28.62)	(47.46)		(33.66)
Physical Proximity to			0.646**	-0.0959
Suspicious Lending			(2.57)	(-1.33)
County FE	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	Yes
Observations	19,753	19,753	19,753	19,753
Num. Counties	2,338	2,338	2,338	2,338
R^2	0.773	0.796	0.693	0.797
Within County R^2	0.555	0.600	0.398	0.601

Panel B. Homophily and Instrumental Variables

Dep. Variable: Flagged By At Least One Primary Measure Per Capita in Zip Code

	(1)	(2)	(3)	(4)
Method:	OLS		IV	
Instrument:			$\geq$ 100 Mi	$\geq$ 500 Mi
Social Proximity to		1.109***	1.182***	1.242***
Suspicious Lending		(28.38)	(14.65)	(13.95)
Demographic Proximity to	0.360***	0.00189		
Suspicious Lending	(2.86)	(0.13)		
County FE	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes
Observations	19,753	19,753	19,753	19,753
Num. Counties	2,338	2,338	2,338	2,338
R^2	0.552	0.773	0.795	0.793
Within County R^2	0.120	0.555	0.597	0.593
First Stage F-stat			18.04	9.76

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table 3. County Cultural Features

This table examines the effect of cultural variables on the relationship between the number of flagged loans per capita and social proximity to suspicious lending at the county level. *Identity Theft* is the average number of identity theft reports to the FTC per capita in each CBSA from 2015 to 2019. *FinCEN SARs* is the average number of suspicious activity reports per capita in each county from 2015 to 2019 (winsorized at the 95th percentile). *Political Corruption* is the number of public corruption convictions in each judicial district from 2010 to 2019 per million residents (as reported by the DOJ). *Religious Affiliation* is the percentage of the county's population with a religious affiliation as of 2020 (as reported by the Association of Religious Data Archives). *Ashley Madison* is the paid Ashley Madison usage rate in each county (as reported by Griffin, Kruger, and Maturana (2019)). Column (4) uses versions of the *Identity Theft* and *FinCEN SARs* variables that are based on data from 2020 and 2021. The controls are log of population density, percentage non-white, log of average household income, poverty rate, pre-pandemic unemployment, percentage with college education in the county, and the shares of the friends of Facebook users in the county who live within 50 and 150 miles of them. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the number of loans in each county. All variables (both independent and dependent) are standardized at the county level to have a mean of 0 and a standard deviation of 1. Fixed effects and controls are as indicated at the bottom of each column. Robust standard errors are clustered at the state level.

Dep. Variable: Flagged By At Least One Primary Measure Per Capita in County

	(1)	(2)	(3)	(4)
Social Proximity to Susp. Lending	0.674***	0.614***	0.586***	0.529***
	(5.96)	(2.99)	(4.95)	(4.19)
Physical Proximity to Susp. Lending		0.143		
		(0.31)		
Identity Theft			0.237**	
			(2.34)	
Identity Theft Contemporaneous				0.246***
				(3.21)
FinCEN SARs			0.0747	
			(1.40)	
FinCEN SARs Contemporaneous				0.110*
				(1.97)
Public Corruption			-0.0406	-0.00597
			(-0.89)	(-0.11)
Religious Affiliation			0.0866**	0.0808**
			(2.64)	(2.32)
Ashley Madison			0.112	0.0913
			(1.14)	(0.88)
Controls	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes
Observations	1,556	1,556	1,556	1,556
R^2	0.802	0.803	0.824	0.826
Within State R^2	0.550	0.552	0.599	0.603

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table 4. Social Capital

This table examines the effect of social capital variables from Chetty et al. (2022a,b) on the relationship between social proximity and flagged PPP loans per capita at the zip code level. The controls are the log of population density, percentage non-white, log of average household income, poverty rate, pre-pandemic unemployment, percentage with college education, the shares of the friends of Facebook users in the zip code who live within 50 and 150 miles of them, and the FinTech market share of PPP loans in the zip code. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the number of PPP loans in each zip code. All variables (both independent and dependent) are standardized at the zip code level to have a mean of 0 and a standard deviation of 1. Zip codes are filtered to those with at least 25 loans. Fixed effects and controls are as indicated at the bottom of each column. Robust standard errors are clustered by county.

Dep. Variable: Flagged By At Least One Primary Measure Per Capita in Zip Code

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Social Proximity to Susp. Lending	1.047***	1.049***	1.042***	1.054***	1.042***	1.043***	1.038***	1.044***
	(40.99)	(40.03)	(41.09)	(41.19)	(38.91)	(39.19)	(36.86)	(39.90)
Economic Connectedness	0.0484**							0.0465*
	(2.07)							(1.86)
Exposure		0.0759**						
		(2.00)						
Friending Bias			-0.0149					
			(-1.52)					
Clustering				0.0867***				0.0776***
				(4.93)				(3.88)
Support Ratio					-0.0204			-0.0165
					(-1.52)			(-1.19)
Volunteering Rate						-0.0209		-0.0349**
						(-1.60)		(-2.40)
Civic Organizations							0.115***	0.113***
							(6.57)	(6.35)
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	16,814	16,814	16,814	18,814	16,814	16,814	16,814	16,814
Num. Counties	2,116	2,116	2,116	2,116	2,116	2,116	2,116	2,116
R^2	0.842	0.842	0.841	0.842	0.841	0.841	0.845	0.847

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table 5. Spread Across Rounds

This table examines the spread of fraud across rounds. In columns (1)-(3), social/physical proximity to suspicious lending and the fraud rate are calculated based on all loans; in columns (4)-(6) and (7)-(9), they are calculated based on only FinTech and traditional loans, respectively. The controls are the log of population density, percentage non-white, log of average household income, poverty rate, pre-pandemic unemployment, percentage with college education, the shares of the friends of Facebook users in the zip code who live within 50 and 150 miles of them, and the FinTech market share of PPP loans in the zip code during 2021. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the number of PPP loans for columns (1)-(3), the number of PPP loans originated by FinTech lenders in columns (4)-(6), and the number of PPP loans originated by traditional lenders in each zip code for columns (7)-(9). All variables (both independent and dependent) are standardized to have a mean of 0 and a standard deviation of 1. Zip codes are filtered to those with at least 25 loans. Fixed effects and controls are as indicated at the bottom of each column. Robust standard errors are clustered by county.

Dep. Variable: 2021 Flagged By At Least One Primary Measure Per Capita in Zip Code

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Based on:	——— All Loans ———			—– FinTech Loans —–			— Traditional Loans —		
Social Proximity to	0.885**		1.002**	0.781***		0.920***	0.0359		0.0146
2020 Suspicious Lending	(2.18)		(2.21)	(3.68)		(3.98)	(1.53)		(0.48)
Physical Proximity to		0.0871**	-0.116*		0.0462	-0.264**		0.0230**	0.0207
2020 Suspicious Lending		(2.28)	(-1.65)		(0.86)	(-2.21)		(2.00)	(1.43)
2020 Flagged	0.472***	0.480***	0.476***	0.588***	0.759***	0.586***	0.810***	0.808***	0.809***
Per Capita	(4.24)	(4.11)	(4.21)	(4.73)	(4.63)	(4.72)	(18.93)	(19.16)	(19.07)
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	19,393	19,393	19,393	19,393	19,393	19,393	19,393	19,393	19,393
Num. Counties	2,308	2,308	2,308	2,308	2,308	2,308	2,308	2,308	2,308
R^2	0.800	0.769	0.802	0.909	0.891	0.912	0.896	0.896	0.896

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table 6. Social Proximity to Lenders

This table examines the effects of social proximity to FinTech loans and loans originated by different lenders/platforms on borrowers using a FinTech lender or specific lender/platform. Social proximity to lending by a lender, or group of lenders, is calculated in the same way as social proximity to suspicious lending, but with the number of flagged loans per capita in each zip code replaced with the number of loans originated by the lender(s) per capita in the zip code. The dependent variables are the number of loans originated by the lender(s) per capita in each zip code. FinTech lenders are those used by Griffin, Kruger, and Mahajan (2023). Womply worked with five lenders to originate loans: Benworth Capital, Capital Plus Financial, DreamSpring, Fountainhead SBF, and Harvest SBF. BlueAcorn worked with two lenders to originate loans: Capital Plus Financial and Prestamos CDFI. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the number of PPP loans in each zip code. All variables (both independent and dependent) are standardized to have a mean of 0 and a standard deviation of 1. Zip codes are filtered to those with at least 25 loans. Fixed effects and control variables are as indicated at the bottom of each column. The controls are the log of population density, percentage non-white, log of average household income, poverty rate, pre-pandemic unemployment, percentage with college education, and the shares of the friends of Facebook users in the zip code who live within 50 and 150 miles of them. Robust standard errors are clustered by county.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dep. Variable:	FinTech Per Capita	Womply Per Capita		BlueAcorn Per Capita		BoA Per Capita	JPMChase Per Capita
Social Proximity to FinTech	0.858*** (13.47)						
Social Proximity to Womply		1.016*** (21.45)	0.838*** (4.12)		-0.115 (-1.30)		
Social Proximity to BlueAcorn			0.192 (0.93)	1.092*** (18.46)	1.204*** (9.32)		
Social Proximity to Bank of America						-0.0438 (-1.04)	
Social Proximity to JPMorgan Chase							0.00676 (0.05)
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Loans Per Capita	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	19,367	19,367	19,367	19,367	19,367	19,367	19,367
Num. Counties	2,320	2,320	2,320	2,320	2,320	2,320	2,320
R^2	0.928	0.897	0.898	0.916	0.916	0.813	0.734

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

For Online Publication

Internet Appendix for:
"Is Fraud Contagious? Social Connections and the Looting of COVID Relief Programs"

A. Details Regarding PPP Fraud Measures

The loan-level indicators of suspicious PPP loans were developed and validated by Griffin, Kruger, and Mahajan (2023). To aid in understanding, we briefly describe each of the primary indicators here. For further details, we refer the reader to Griffin, Kruger, and Mahajan (2023).

Business Registry Flag

Businesses organized as corporations, S-corporations, and LLCs are required to file an article of incorporation or LLC filing with a state, either as a domestic company in their home state or as a foreign company in another state. Further, the SBA required businesses to be "in operation on February 15, 2020... [and] not permanently closed." Based on these requirements, the following conditions are checked for all corporation, S-corporation, and LLC borrowers:

(i). Is there a matching business in the business registry data? ("Missing Business")

(ii). Was the business dissolved and inactive before being approved for a PPP loan? ("Dissolved Business")

(iii). Is the earliest incorporation or initial filing date for the business after February 15, 2020? ("Late Incorporation/Filing")

These three subflags are combined to form an overall business registry flag.

Multiple Loans Flag

While a business owner may have multiple businesses registered to the same address, the presence of multiple loans at an individual residential address during the same draw is also a potential sign of fictitious operations. This flag identifies residential (i.e., nonbusiness, noncentral) standardized addresses with three or more loans within the same draw. Note that the flag uses a cutoff of three loans instead of two loans in order to be more conservative.

High Implied Compensation Flag

PPP loan size is limited to 2.5 times a business's average monthly payroll expenses, including up to $100,000 in annual compensation per employee. PPP loan applications report how many employees the business has based on the same time period used to calculate average payroll expenses (2019 in most cases). Based on loan size and number of reported employees, one can impute implied average annual compensation. Loans for which the implied compensation per job reported is more than three times the industry-CBSA average compensation/receipts are flagged by this indicator.

EIDL Advance Jobs > PPP Jobs Flag

Concurrently with the PPP, the SBA provided businesses and individuals with the ability to receive a forgivable EIDL advance of up to $10,000. For all EIDL advances issued in 2020, the advance amount was calculated as $1,000 per employee (up to the $10,000 maximum). Thus, there was an incentive for borrowers to inflate the number of jobs reported on their EIDL applications. This flag identifies borrowers that appear to have manipulated the number of employees reported on their EIDL applications to exploit this incentive. To make it less likely that differences are driven by reporting or timing differences in the number of employees reported, only loans where the individual claimed three or more jobs extra on their EIDL applicant as compared to their PPP application are flagged by this indicator.

Validation of PPP Fraud Measures

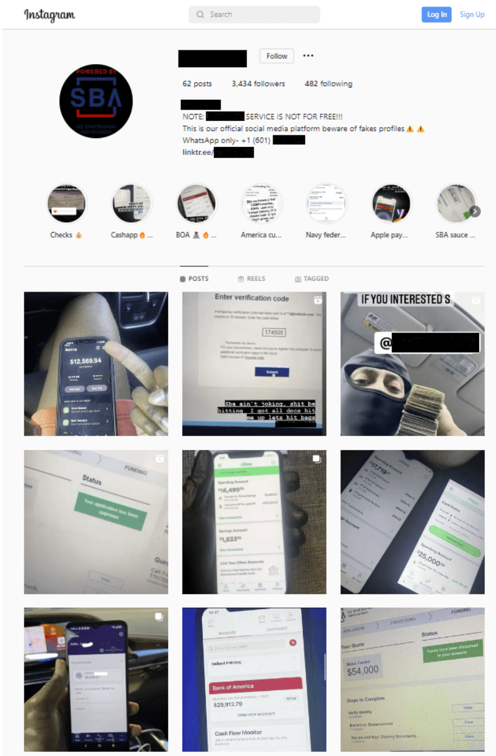
See Griffin, Kruger, and Mahajan (2023) for validation of the fraud measures including secondary measures of fraud, independent external measures. A Congressional investigation into PPP fraud (see Congressional report here), also validates the findings of Griffin, Kruger, and Mahajan (2023), particularly with respect to the high fraud rates they find for many FinTech lenders.

B. Social Media Activity

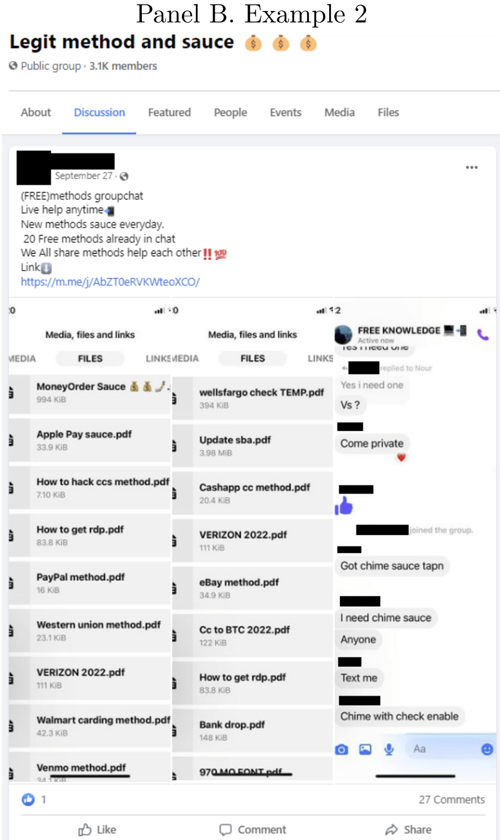
Exhibit IA.1. Examples of Suspicious Activity Related to Pandemic Relief Programs

This exhibit shows some examples of social media activity related to pandemic relief programs. Any names or identifying information (such as phone numbers) are redacted.

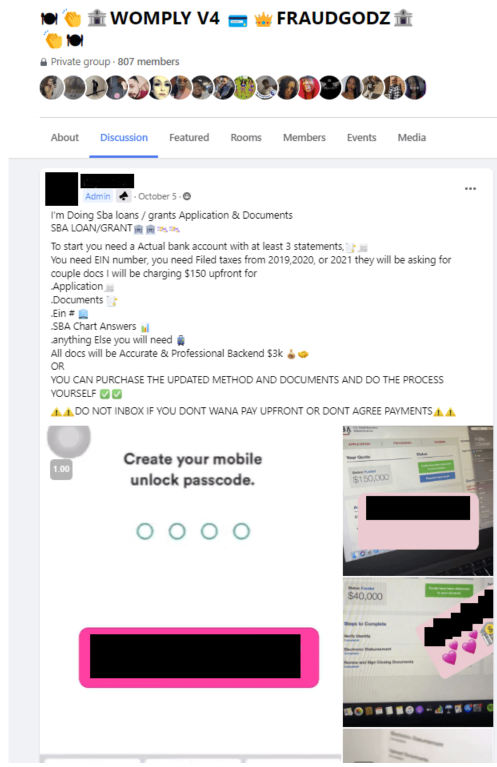
Panel A. Example 1



Panel B. Example 2

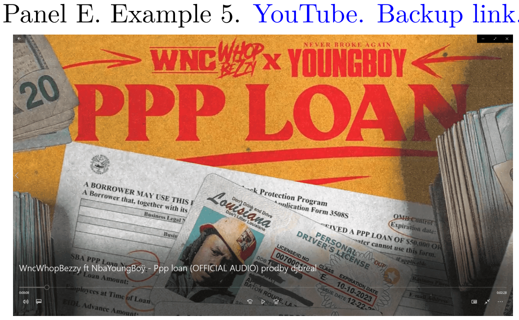


Panel C. Example 3



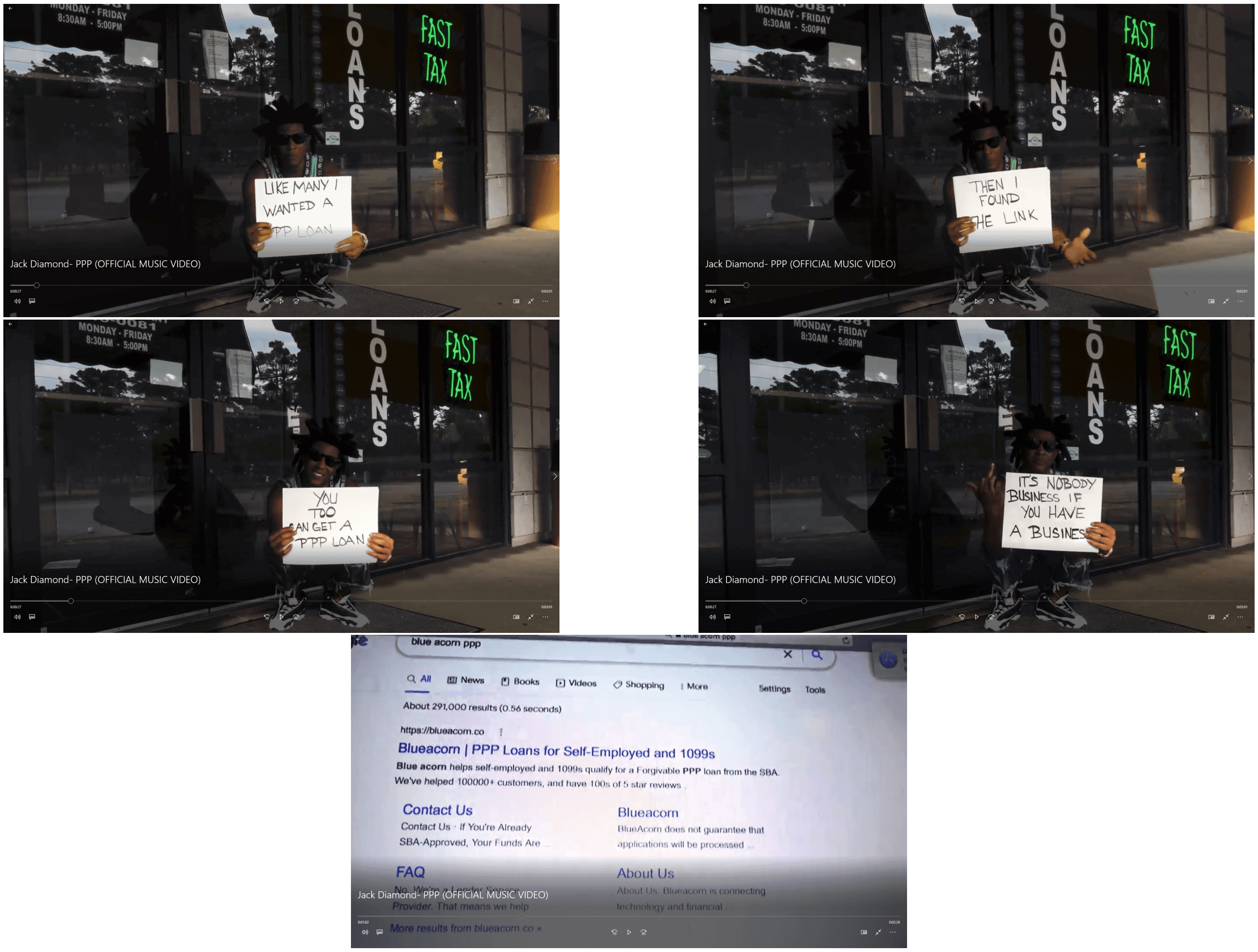
Panel D. Example 4

Panel E. Example 5. YouTube. Backup link.

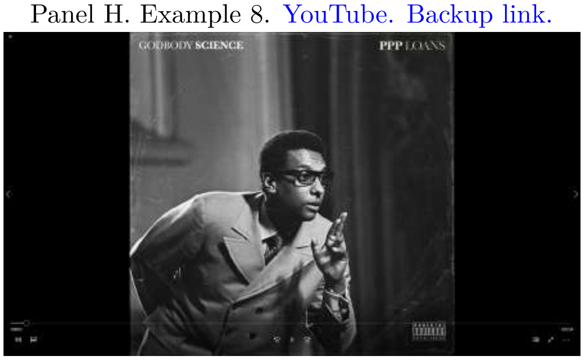


Panel F. Example 6

Panel G. Example 7. YouTube. Backup link.



Panel H. Example 8. YouTube. Backup link.



Panel I. Example 9

[redacted] • 2 yr. ago

so the thing is, the amount of money they are taking out of the pool is nothing. The $20,000 that they are getting is pennies. We get scammed from the government every day of our lives, society is TOO SLEEP to realize it. You have to think about this, we are fresh off of a pandemic, slow economy. At this point the government is just going money out. they don't care who gets it and what its for, just don't get too greedy or too flashy!

1 Reply Share

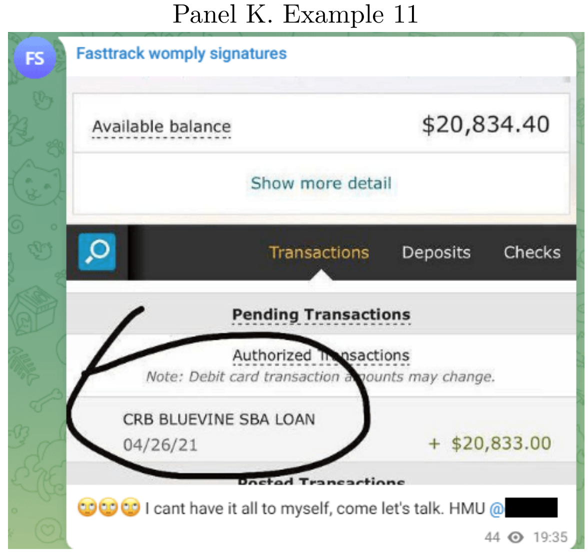
5 more replies

Panel J. Example 10

[redacted]

Update Pros (Fullz info) October 09
CVV Fresh with balance high
My SBA SAUCE IS GUCCI,
THERE'S NO WAY YOU APPLY USING MY METHOD AND DON'T GET PAID.
APPROVED AMOUNT FUNDED TO YOUR BANK ACCOUNT DIRECTLY
Also available......
Womply PPP sauce
Bluevine updated method
Willa sauce

Panel K. Example 11



Panel L. Example 12

[redacted] • 9 mo. ago

Literally everyone I know done it and they never got caught. r[redacted]as keep saying they coming for everyone just wait but it's literally impossible to prosecute every single person who took one out. they don't have the resources for that

13 Reply Share

Panel M. Example 13

Post

[redacted]
@[redacted]

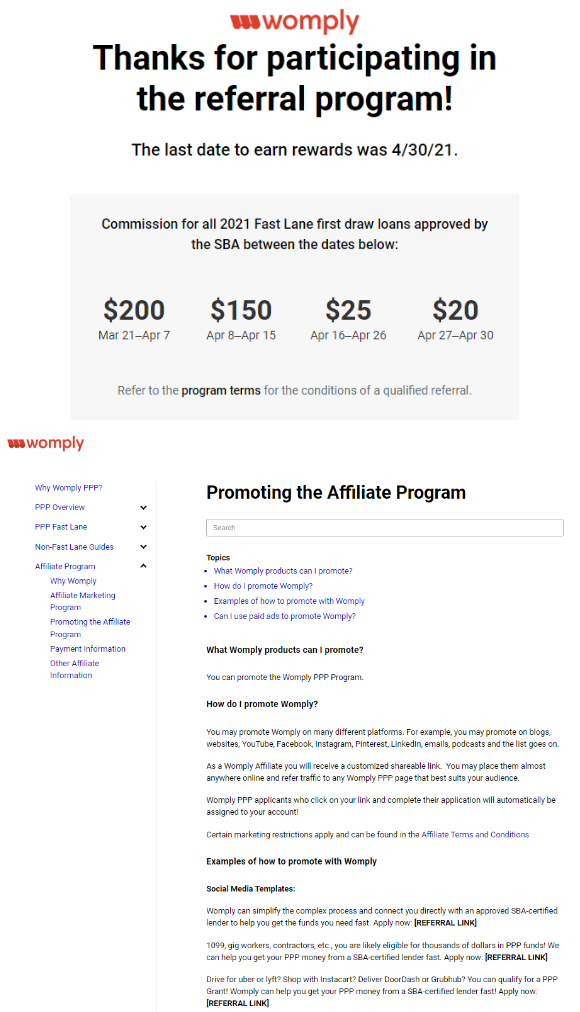
Don't talk money too me #inscamwetrust #swipe #edd #methods #loans #ppp #sauce #fire

12:36 PM · Jul 27, 2021

Exhibit IA.2. Examples of Marketing by FinTechs

This exhibit shows some examples of marketing activity by FinTechs, in particular by Womply and BlueAcorn, related to the PPP. Any names or identifying information (such as phone numbers) are redacted.

Panel A. Womply Affiliate Program



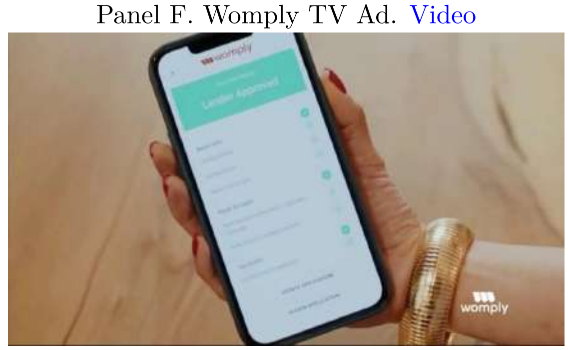
Panel B. Womply Bus Ad. Source: NY Times

Panel C. NASCAR Sponsorship by Womply. NASCAR

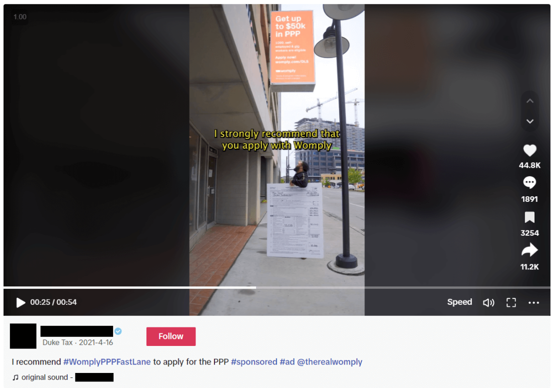
Panel D. Womply Facebook Ad

Panel E. Womply Billboard. Source: LA Times

Panel F. Womply TV Ad. Video



Panel G. Womply Sponsored TikTok. Video



Panel H. Womply LinkedIn Post

Panel I. BlueAcorn Affiliate Marketing, Example 1

Panel J. BlueAcorn Affiliate Marketing, Example 2

Panel K. BlueAcorn Facebook Ad, Example 1

Panel L. BlueAcorn Facebook Ad, Example 3

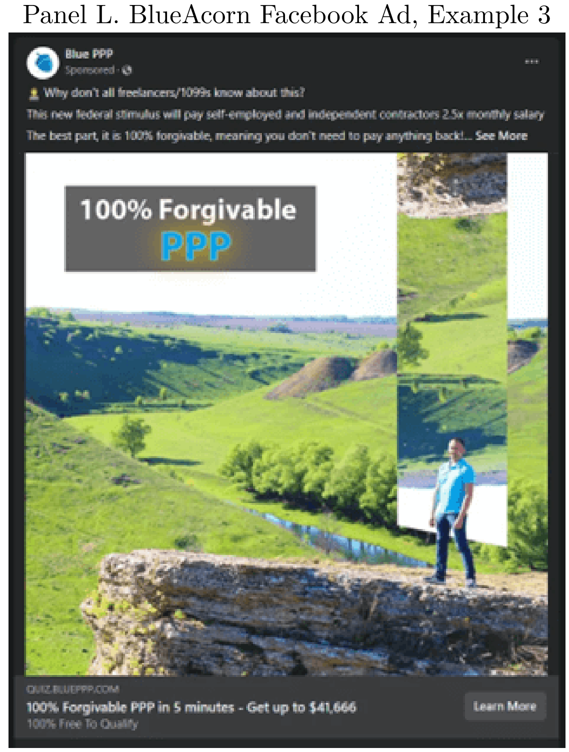


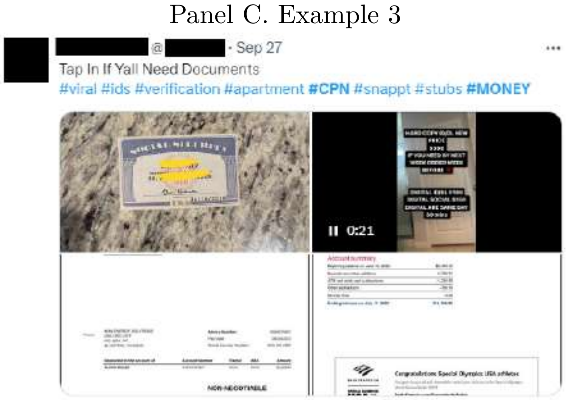
Exhibit IA.3. Examples of Suspicious Activity in 2022-23

This exhibit shows some examples of social media posts in 2022-23 related to suspicious financial activity. Any names or identifying information (such as phone numbers) are redacted.

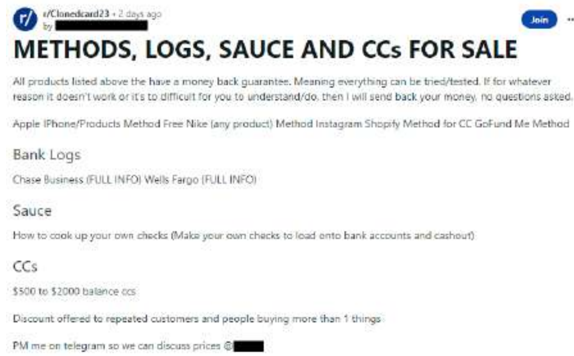
Panel A. Example 1

Panel B. Example 2

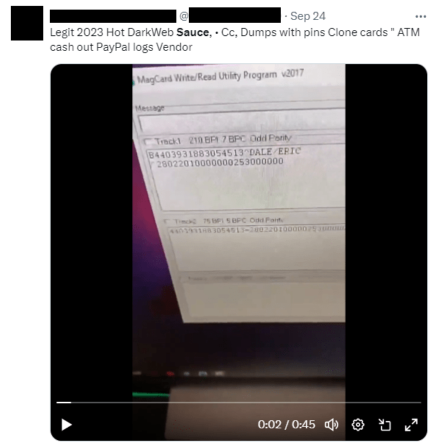
Panel C. Example 3



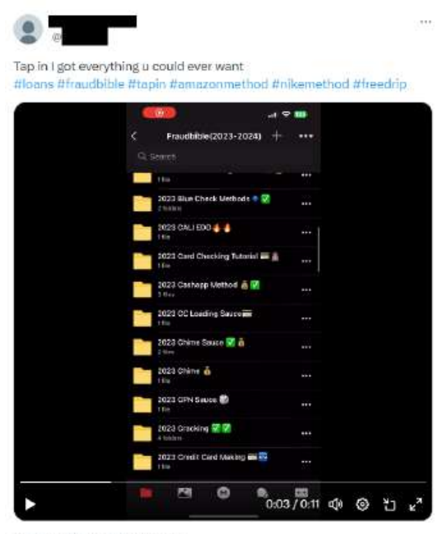
Panel D. Example 4



Panel E. Example 5



Panel F. Example 6

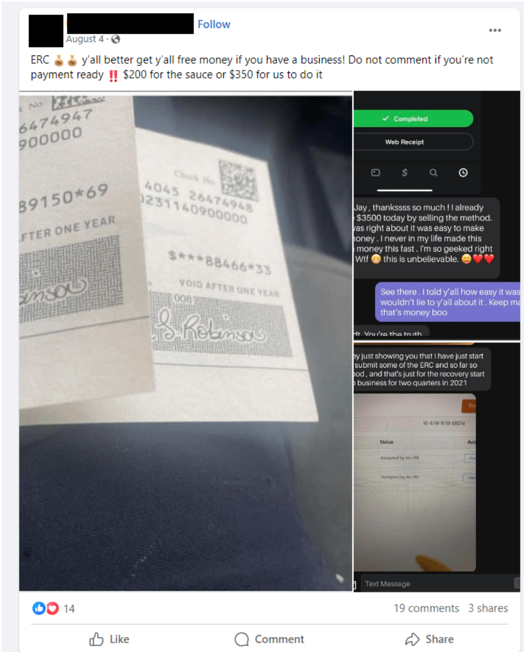


Panel G. Example 7

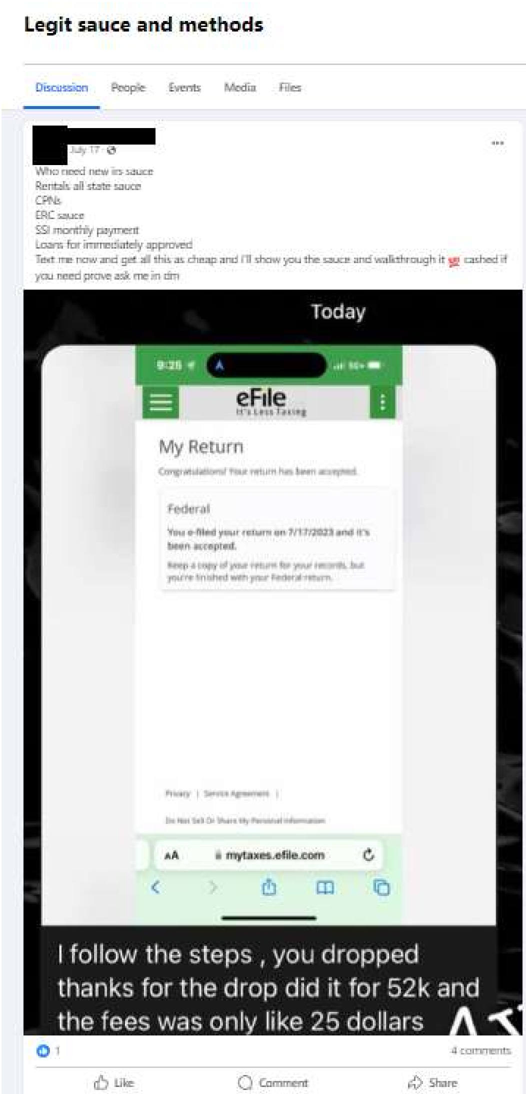
Exhibit IA.4. Examples of Activity Related to the Employee Retention Tax Credit (ERC)

This exhibit shows some examples of social media posts in 2022-23 related to the Employee Retention Tax Credit (ERC). Any names or identifying information (such as phone numbers) are redacted.

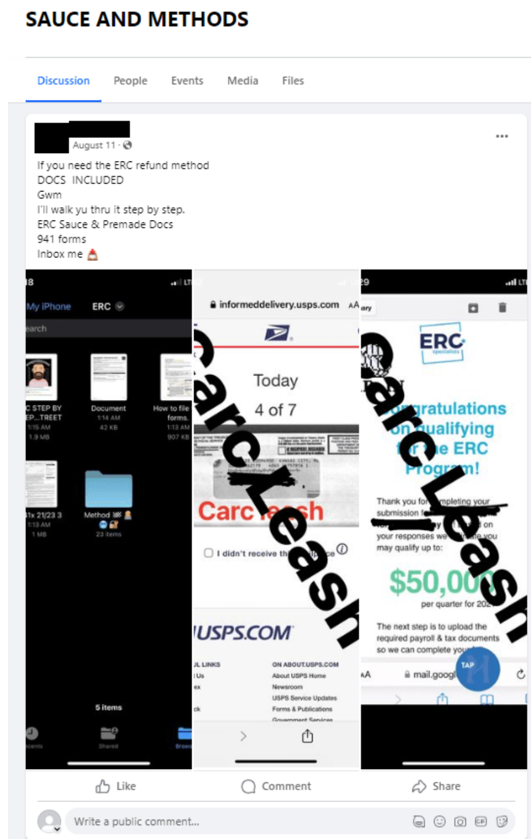
Panel A. Example 1



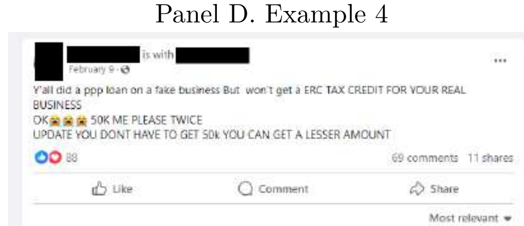
Panel B. Example 2



Panel C. Example 3



Panel D. Example 4



Panel E. Example 5

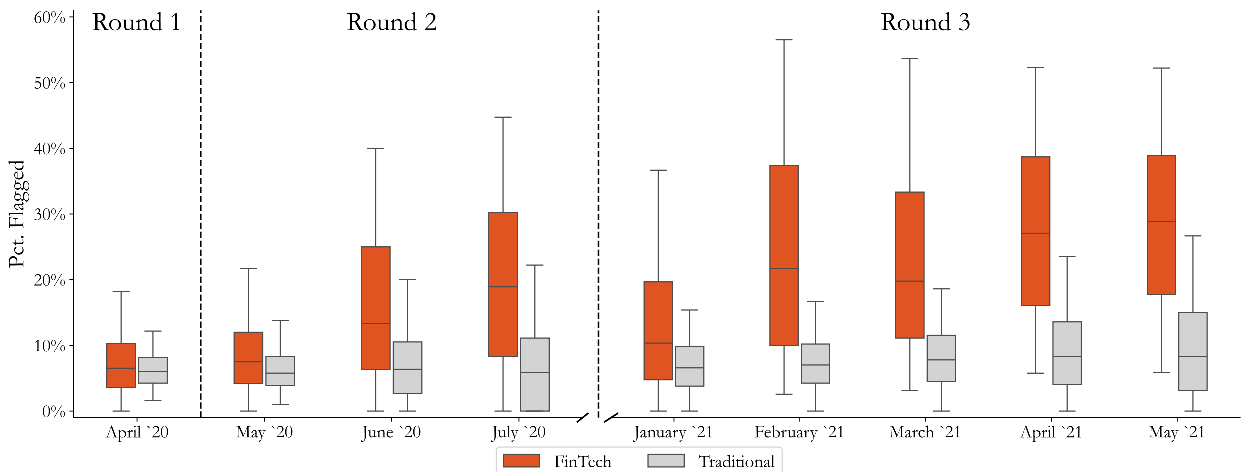
Panel F. Example 6

C. Additional Figures and Tables

Figure IA.1. Geography

This figure shows additional results related to geographic variation (extending Figure 1). Panel A replicates Figure 1, Panel C with data at the zip code level, instead of at the county level. For each month, the distributions of flag rates across zip codes for loans originated by traditional and FinTech lenders are shown by the grey and red boxplots, respectively. The boxplots are weighted by the number of loans in the given zip code-lender type-month cell. Zip code-lender type-month cells with fewer than 10 loans are excluded. Panel B shows the percentage of loans flagged in rounds 1 and 2 is shown on the horizontal axis and in round 3 on the vertical axis. The left subpanel is based on all loans, the center on FinTech loans, and the right on traditional loans. Zip codes with at least 25 loans in rounds 1 and 2 and, for the FinTech and traditional subpanels, 25 loans by the given lender type in rounds 1 and 2 are shown. The black line is a 45-degree line and the correlation is shown at the bottom of each panel. Each dot is a zip code. The size of each dot corresponds to the number of loans in the zip code by the given lender type. The color of each dot is based on the color bar to the right of the panel and corresponds to the growth/decline in lending in the zip code by the given lender type.

Panel A. Distribution of Zip Code Level Fraud Rates by Month and Lender Type



Panel B. Growth and Persistence in Flag Rates Across Rounds, by Lender Type

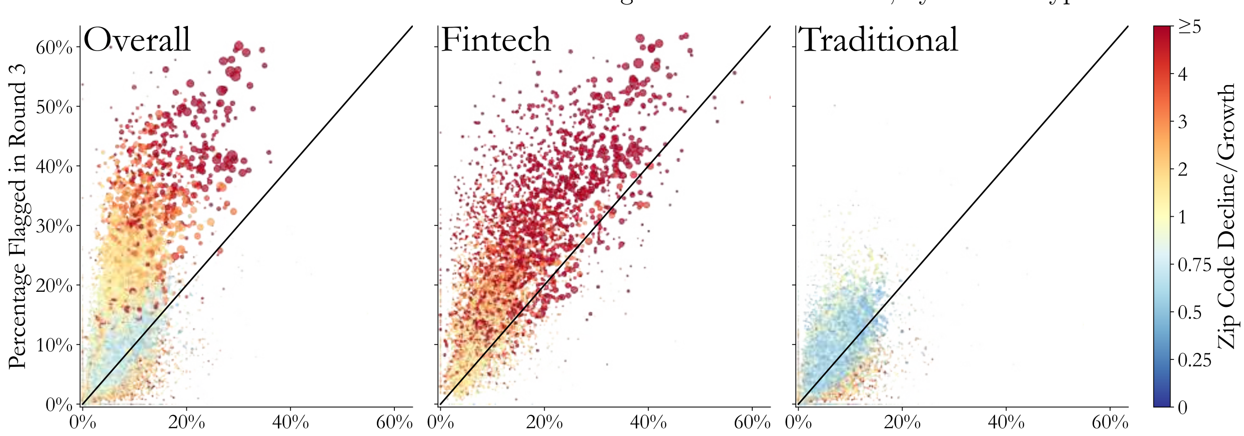
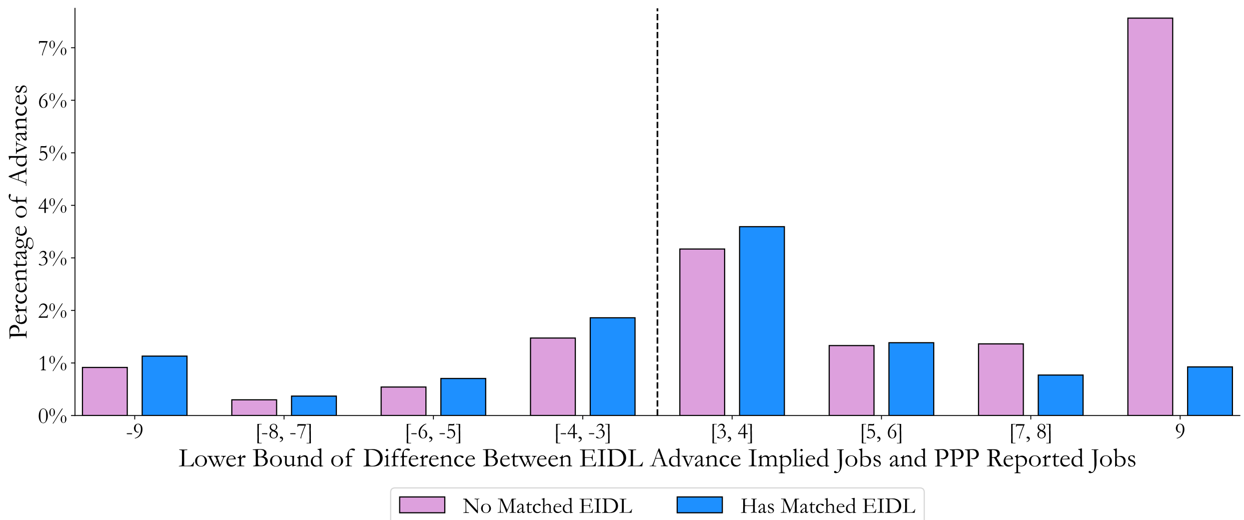


Figure IA.2. EIDL Advance Program

This figure shows additional details regarding the EIDL Advance program. Panel A shows the percentage of EIDL Advances receipts with varying degrees of job mismatches between the jobs implied by their advance amount and the number of jobs reported on their PPP application, by whether the EIDL Advance has a corresponding EIDL loan component. Panel B replicates Figure 2, Panel A after demeaning using county fixed effects. The horizontal axis is based on the percentage of PPP loans that are flagged by at least one of the business registry, multiple loans, or high compensation flags. The vertical axis is based on the percentage of EIDL Advances without a corresponding loan component of the EIDL. Each dot is a zip code. Dots are colored based on the percentage of EIDL Advances receipts with more than three additional jobs implied by their grant amount than they reported on their PPP application. Dots are sized based on the number of EIDL Advances in each zip code. The dashed lines are linear fits and correlations are shown at the bottom left corner of each panel.

Panel A. Job Mismatch Between PPP and EIDL Advance, by Whether Advance Has Corresponding Loan



Panel B. Correlation Between PPP and EIDL Advance Program, Within County

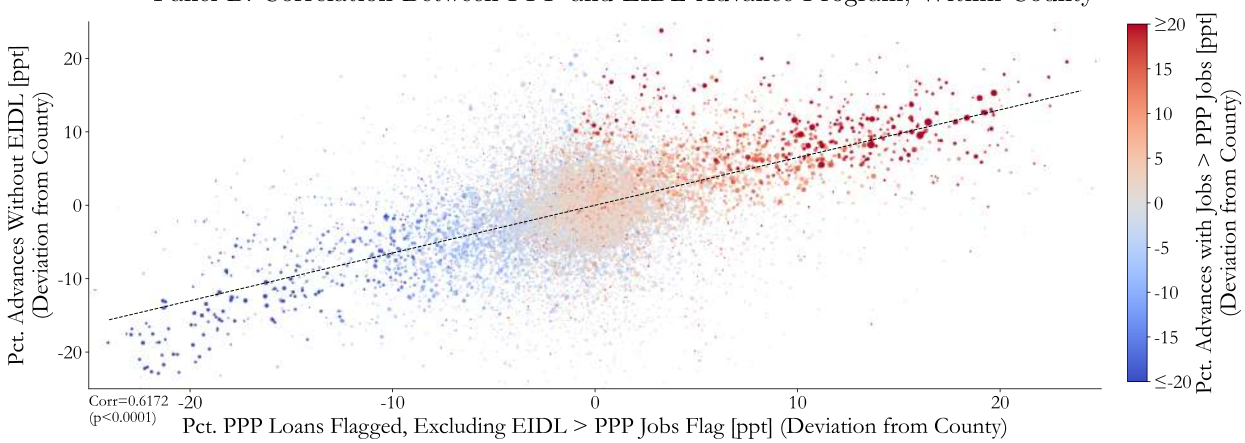
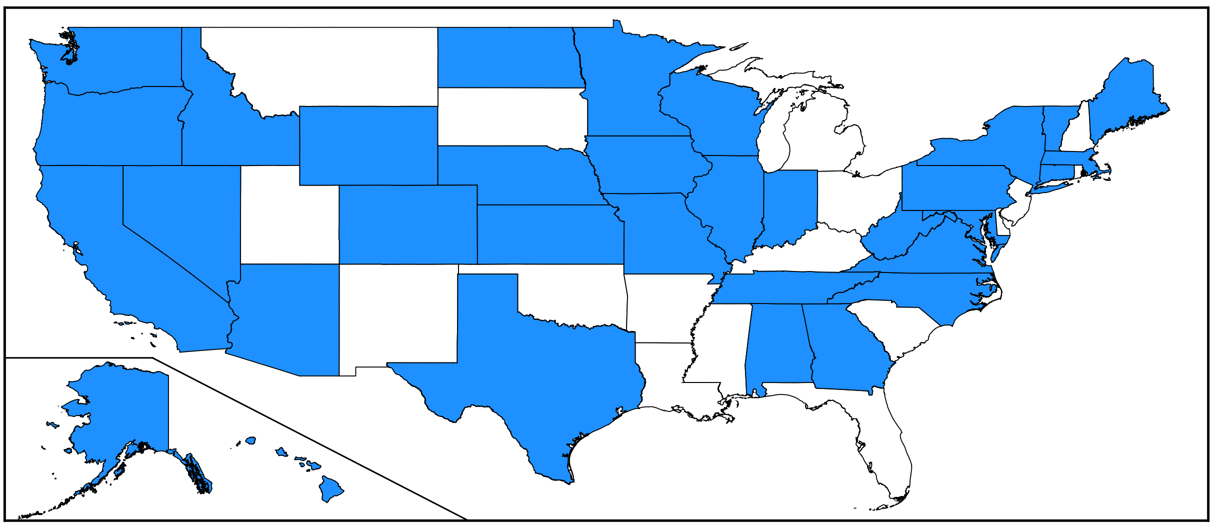


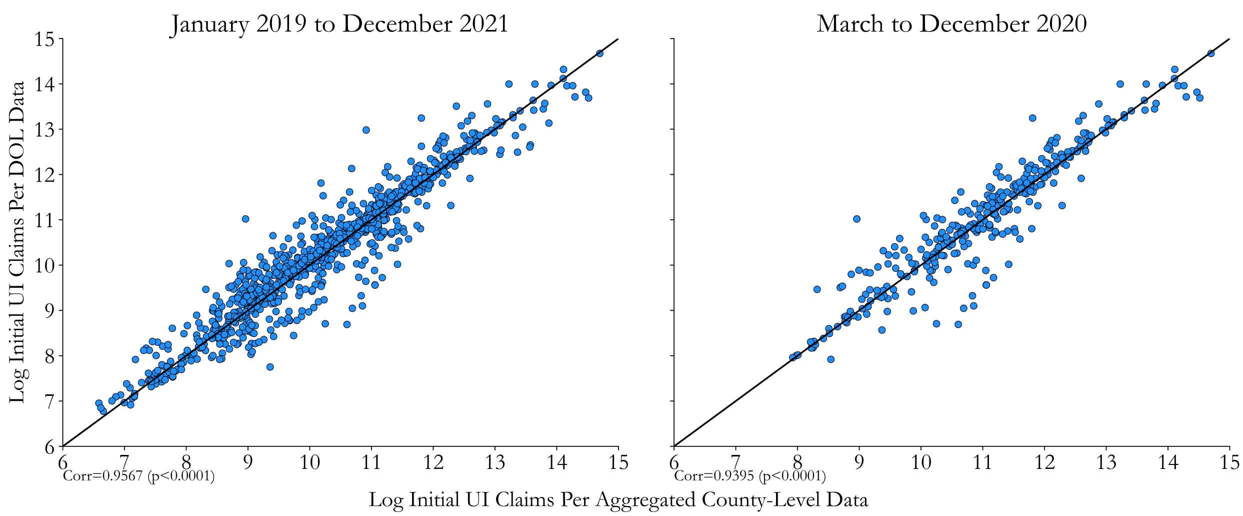
Figure IA.3. Correlation with Unemployment Insurance, Additional Information

This figure provides additional information regarding the unemployment insurance analysis. Panel A shows the states for which we have UI data. Panel B aggregates the county-month-level data to the state-month-level and compares it to state-month-level data reported by the Department of Labor (DOL). The solid line is a 45-degree line and the correlation is shown in the bottom left corner of each subpanel. Panel C shows the effect of PPP loans per capita based on the same regression model used for the top subpanel of Figure 2, Panel B. Panel D replicates the smaller subpanels of Figure 2, Panel B for consumer spending and small business revenue, based on data from the Economic Tracker by Opportunity Insights (described in Chetty et al. (2023)). Panel E shows scatterplots corresponding to Columns (1) and (4) of Table IA.III, Panel A. The dashed line is a linear fit and the correlation is shown in the bottom left corner of each subpanel.

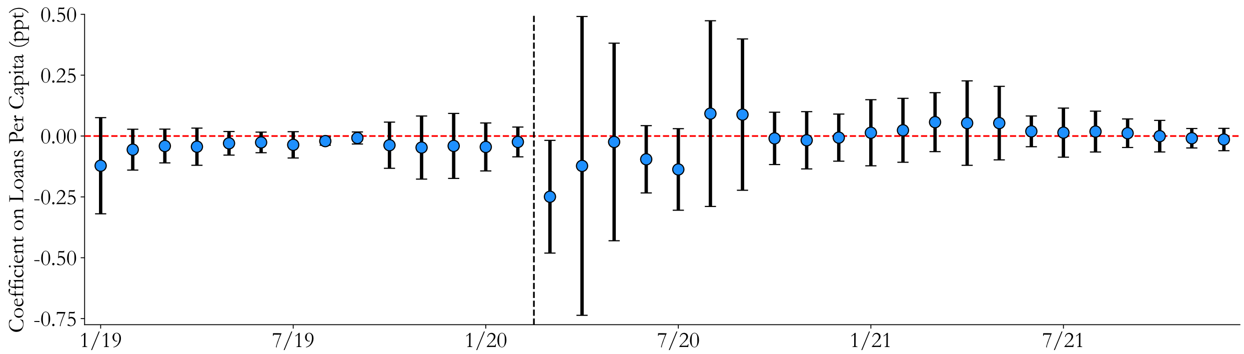
Panel A. States Used in Analysis



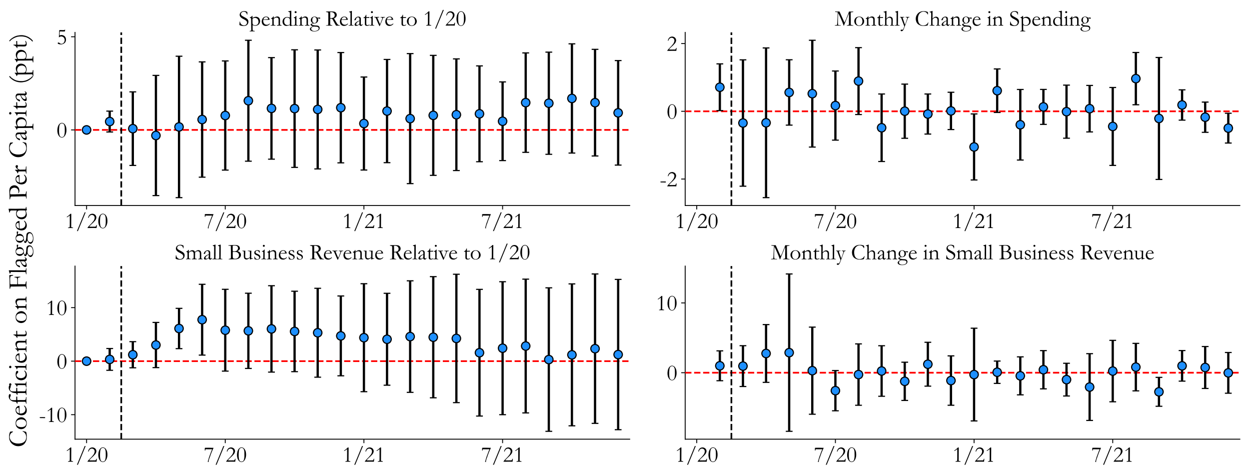
Panel B. Data Validation with DOL State-Level Data



Panel C. Effect of Loans Per Capita on Initial UI Claims



Panel D. Additional Placebo Tests



Panel E. Scatterplots

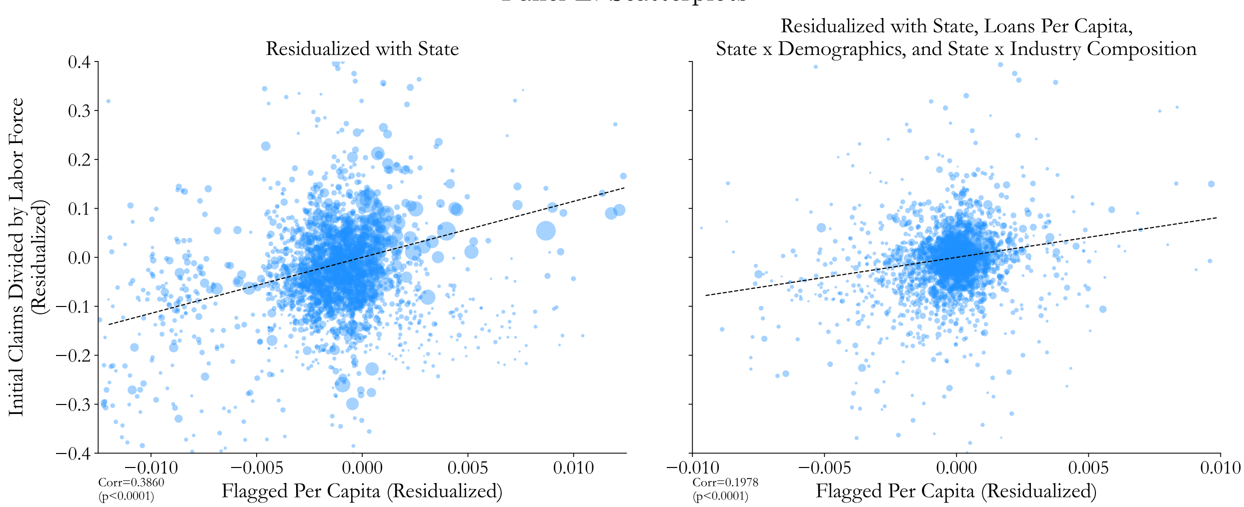


Figure IA.4. Social and Physical Proximity, Separately for FinTech and Traditional

This figure replicates Panel A of Figure 4 using social/physical proximity to suspicious lending and flagged loans per capita calculated separately for FinTech and traditional loans. Proximity measures are standardized to have a mean of 0 and a standard deviation of 1. Each dot represents a zip code, dots are sized based on the number of loans by the lender type in the zip code, and the coloring is based on the flagged PPP loans per capita in the zip code for each lender type (based on the color bar). Zip codes are filtered to those with at least 25 loans. The dashed lines are linear fits and correlations are shown in the bottom left corner of each subpanel.

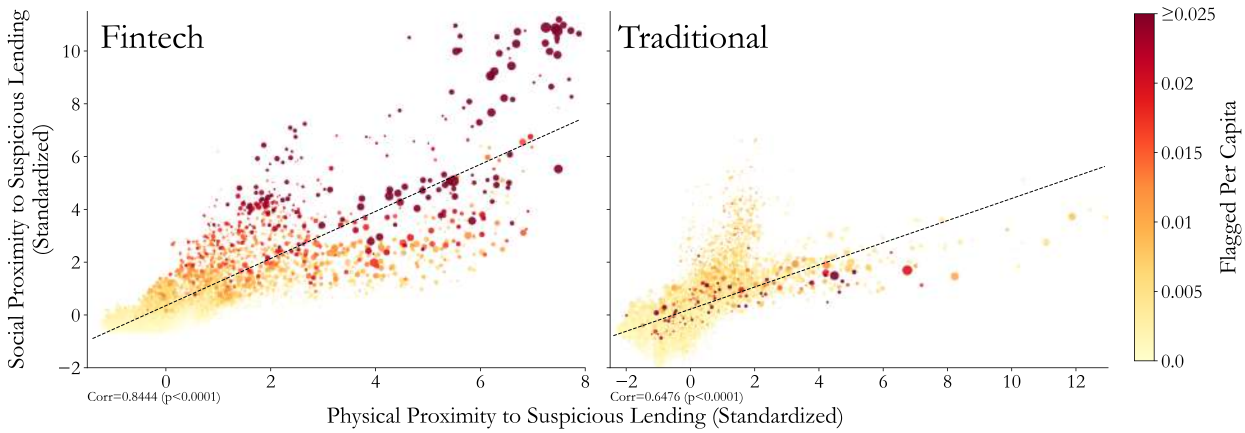
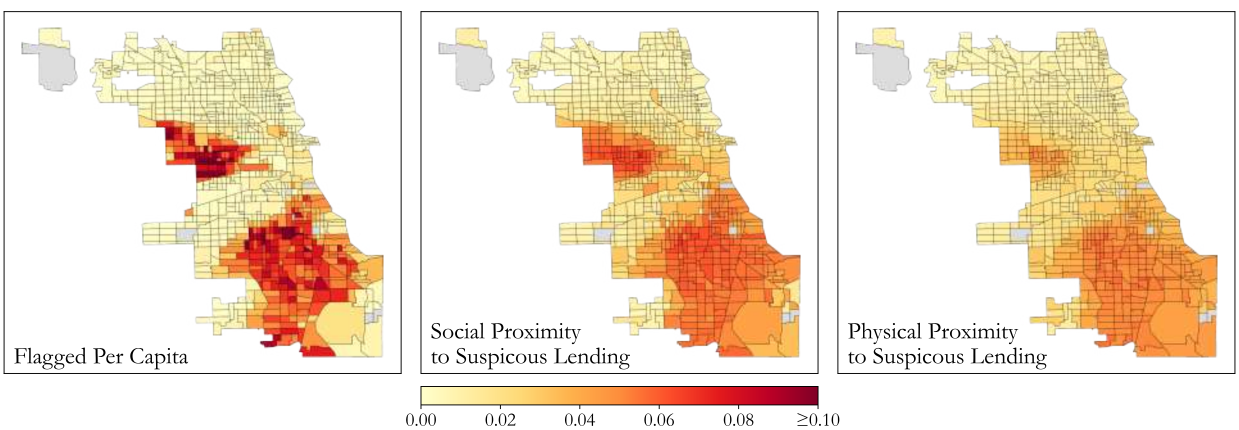


Figure IA.5. Social and Physical Proximity to Suspicious Lending, Based on Chicago Taxi and Rideshare Data

This figure shows the relationship between social proximity to suspicious lending based on an alternative measure and flagged PPP loans per capita in Chicago. Specifically, we use taxi trip data from 2013 to 2019 and rideshare tip data from 2018 to 2019 (see here and here). Analogous to the Social Connectedness Index from Bailey et al. (2018a, 2020), we calculate the social connectedness between Census Tracts as $SCI^{Taxi}_{i,j} = \frac{Trips_{i,j}}{\sum_k Trips_{i,k} \times \sum_k Trips_{k,j}}$ where $Trips_{i,j}$ is the number of taxi and rideshare trips between Census Tracts i and j. Then, $SCI^{Taxi}_{i,j}$ is used to calculate social proximity to suspicious lending in the same way as done previously using the Social Connectedness Index from Bailey et al. (2018a, 2020). Panel A shows heatmaps of flagged per capita, social proximity to suspicious lending, and physical proximity to suspicious lending at the Census Tract level. Coloring is based on the color bar below the heatmaps and Census Tracts with fewer than 25 loans are shown in grey. Panel B shows binscatters based on data at the Census Tract level and weighted by the number of loans in each Census Tract. Both subpanels of Panel B control for population density, percentage non-white, median income, poverty rate, pre-pandemic unemployment, percentage with college education, and the FinTech market share of PPP loans in the Census Tract. The left (right) subpanel additionally controls for physical (social) proximity to suspicious lending and shows the relationship between flagged PPP loans per capita and social (physical) proximity to suspicious lending. Census Tracts are filtered to those with at least 25 loans and the dashed lines are linear fits.

Panel A. Heatmaps



Panel B. Binscatters

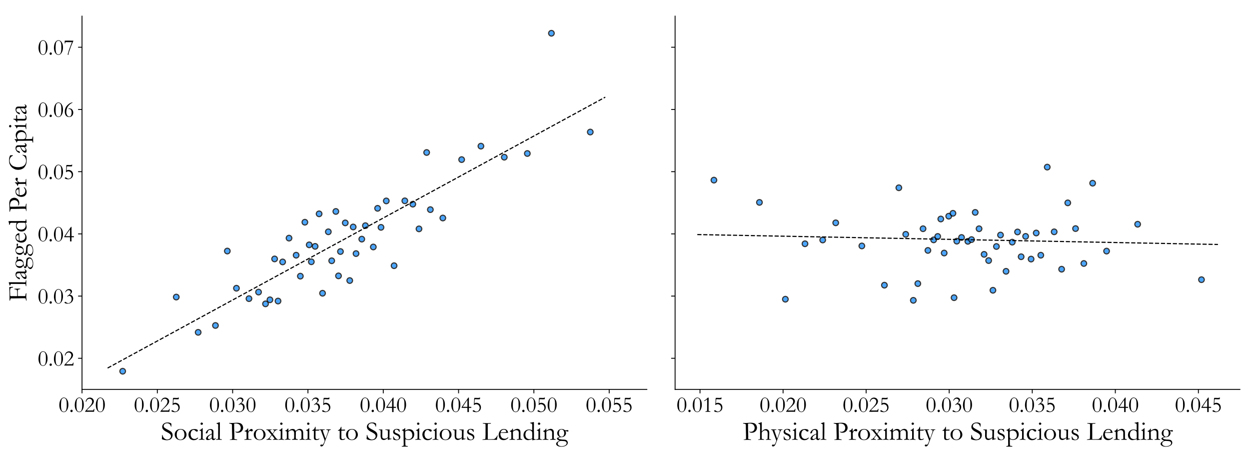
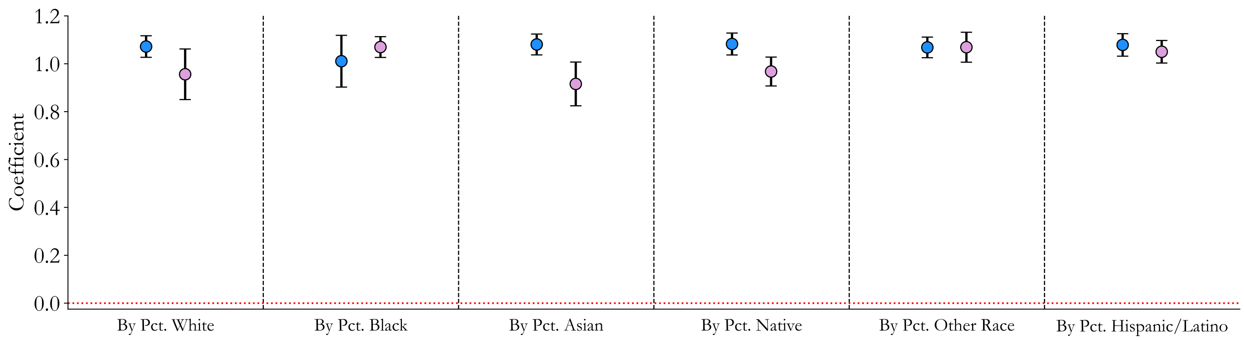


Figure IA.6. Heterogeneity in Effect of Social and Physical Proximity, by Race and Ethnicity

This figure shows heterogeneity in the relationship between social/physical proximity to suspicious lending and flagged per capita across race and ethnicity splits. Panel A shows heterogeneity in the effect of social proximity and Panel B in physical proximity. The regressions include county fixed effects and control for log population density, percentage non-white, the log of average household income, poverty rate, pre-pandemic unemployment, percentage with college education, the share of Facebook friends within 50 and 150 miles, and the FinTech market share of PPP loans in the zip code. The splits are based on the median value of the demographic across all zip codes. The error bars represent 95% confidence intervals based on robust standard errors that are clustered by county. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the number of PPP loans in each zip code. Zip codes are filtered to those with at least 25 loans.

Panel A. Social Proximity



Panel B. Physical Proximity

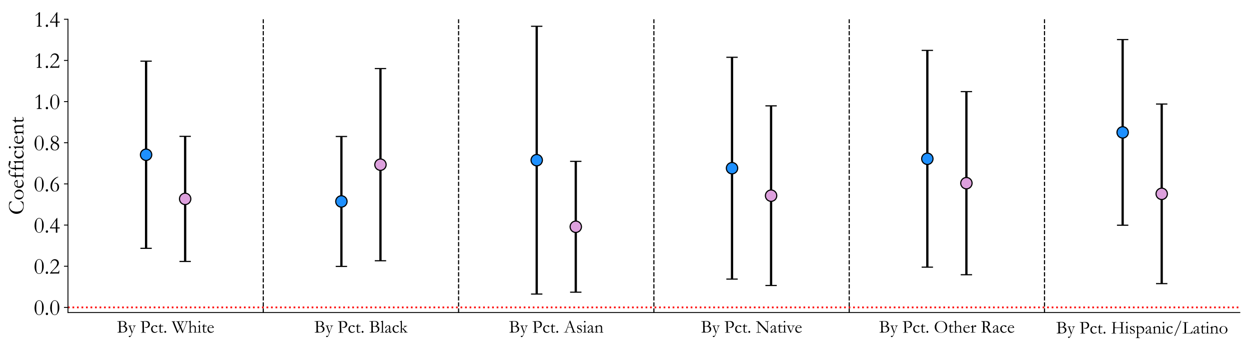


Figure IA.7. Social Proximity, Alternative Measures

This figure replicates the left subpanel of Figure 4, Panel B using alternative measures of social proximity to suspicious lending. All of the subpanels include county fixed effects and control for physical proximity to suspicious lending, log population density, percentage non-white, the log of average household income, poverty rate, pre-pandemic unemployment, percentage with college education, the share of Facebook friends within 50 and 150 miles, and the FinTech market share of PPP loans in the zip code. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the number of PPP loans in each zip code. Zip codes are filtered to those with at least 25 loans.

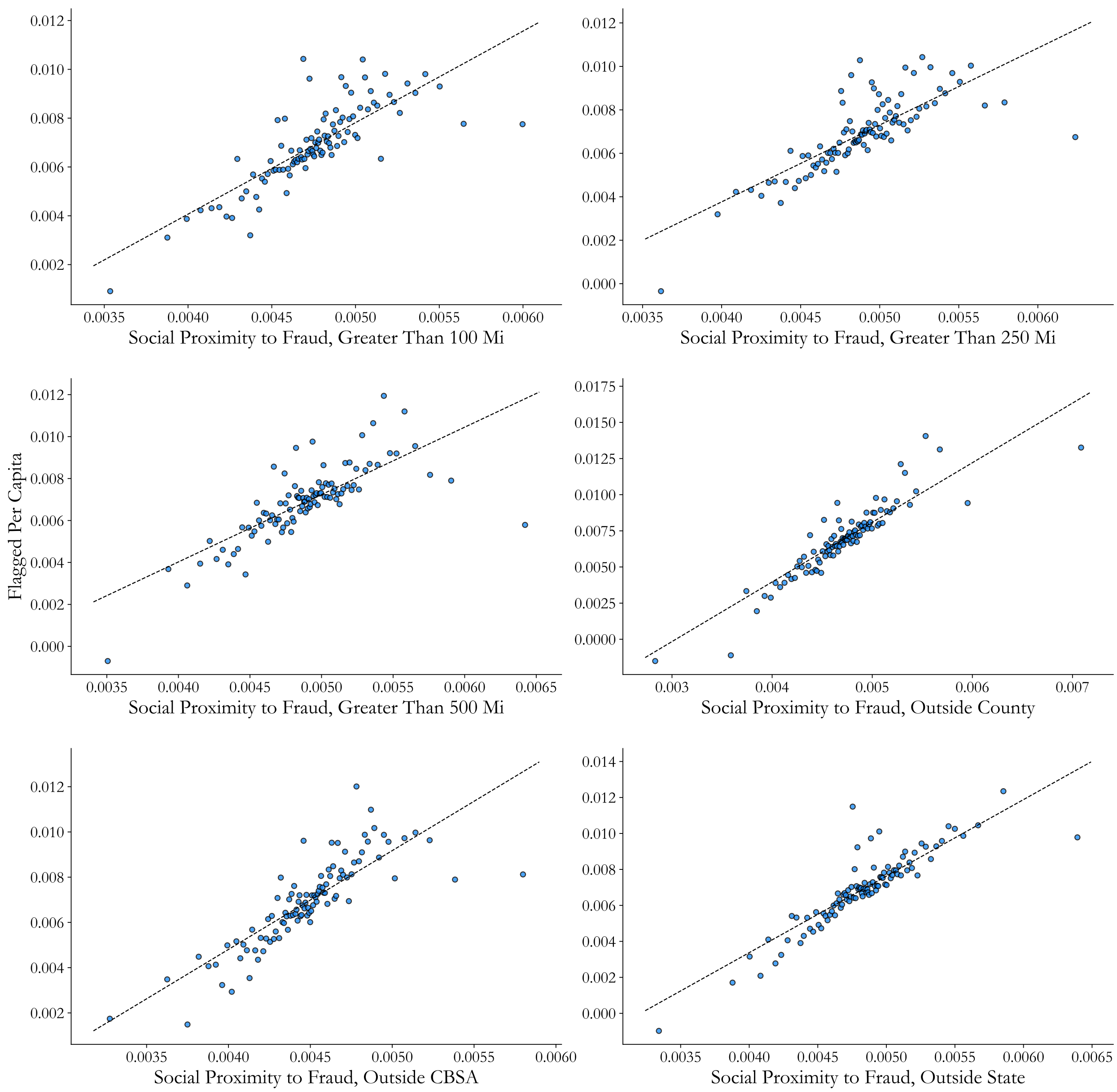
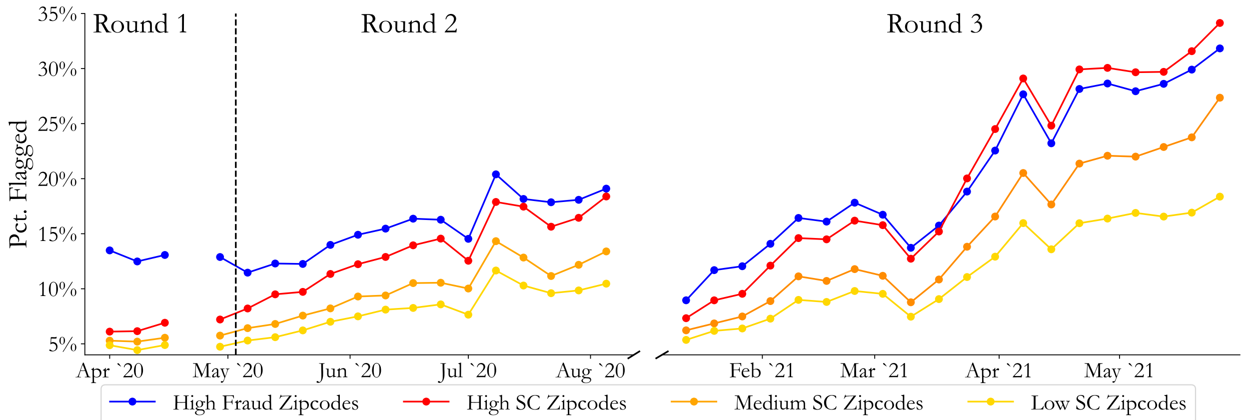


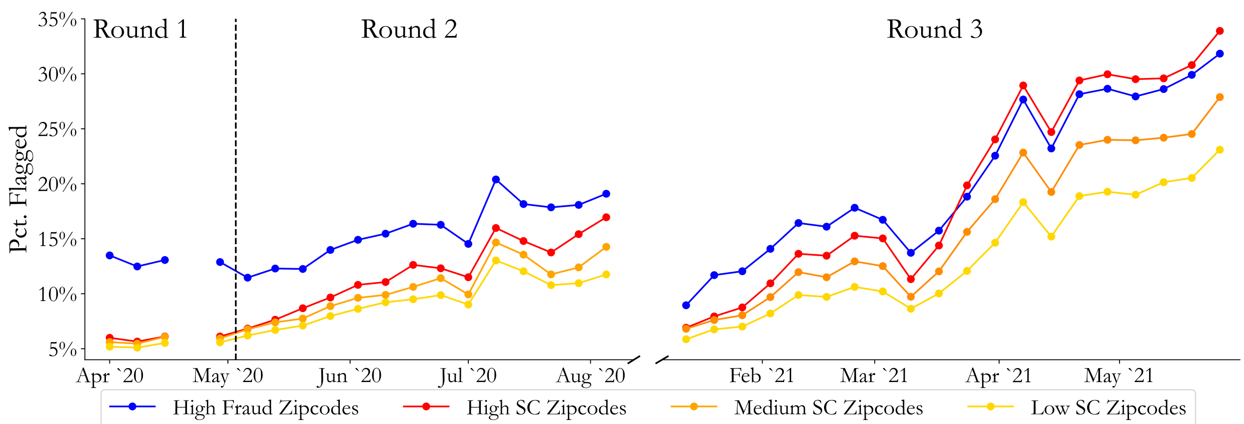
Figure IA.8. Spread of Fraud Over Time

This figure shows additional versions of Figure 6, Panel B. Zip codes are sorted based on their flag rates during the first month of the program and the top zip codes that collectively represent ten percent of lending during the first month are the initial high fraud zip codes. The remaining zip codes are split into terciles, across the entire country in Panel A and within counties in Panels B and C, based on their social connectedness to the initial high fraud zip codes. Specifically, for each zip code i, $Social\ Connectedness_{i,Initial\ High\ Fraud} = \sum_{j \in Initial\ High\ Fraud} Population_j \times SocialConnectedness_{i,j}$. In Panel C, zip codes in the top tercile of social connectedness to the initial high fraud zip codes are matched to a zip code in the same county and in the bottom tercile based fraud rates during the first month of the PPP. The dashed lines represent 95% confidence intervals based on standard errors clustered at the zip code level. The p-value for the difference in percentage flagged during the first month of the program in zip codes in the top tercile of social connectedness and the matched zip codes in the bottom tercile is noted on the figure.

Panel A. Without Matching



Panel B. Within County Splits



Panel C. Within County Splits and Matching

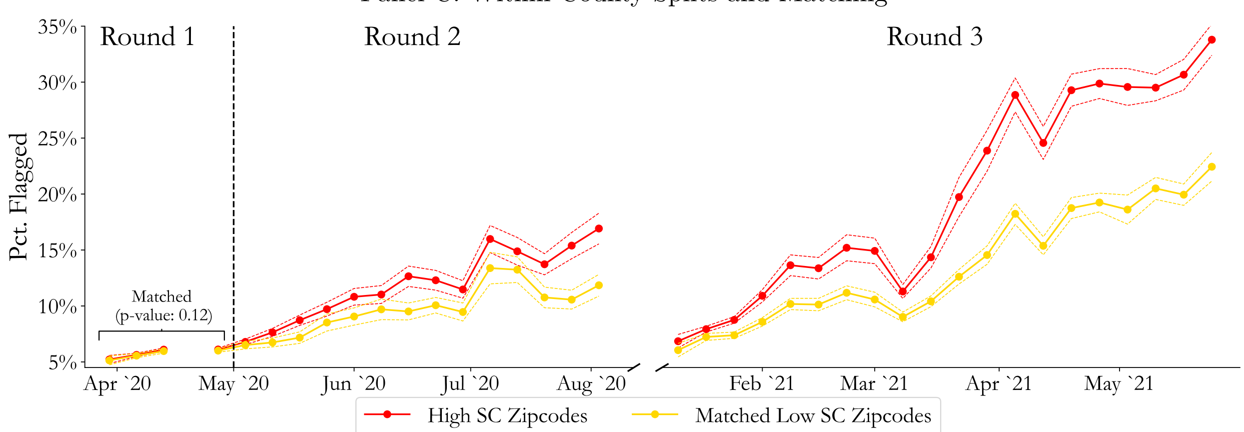
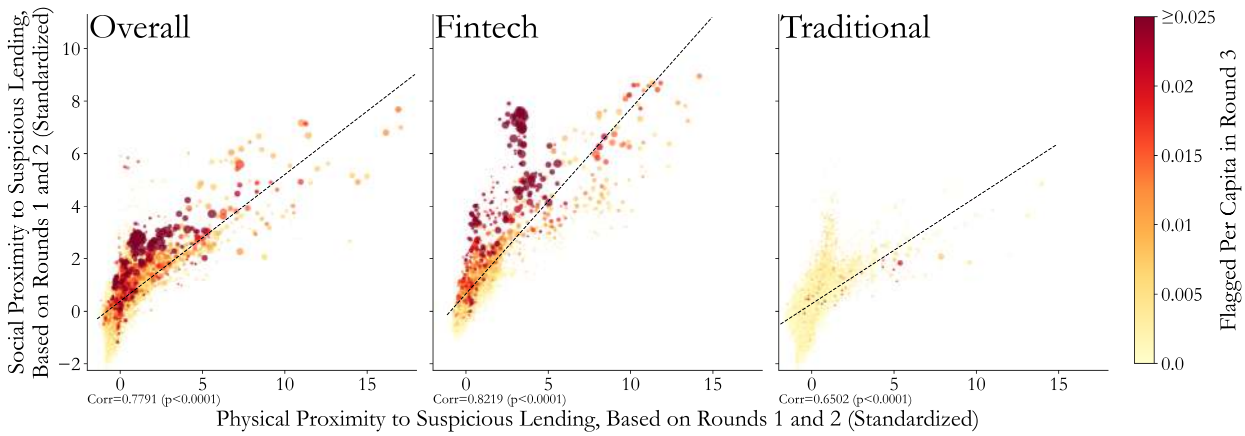


Figure IA.9. Social and Physical Proximity, Rounds 1 and 2

This figure shows the relationship between social/physical proximity to suspicious lending based on rounds 1 and 2 loans and flagged per capita in round 3 (Panel A) and the growth in lending from rounds 1 and 2 to round 3 (Panel B). The proximity measures are standardized to have a mean of 0 and a standard deviation of 1. The left subpanel in each panel is based on all loans, the center subpanel is based on FinTech loans, and the right subpanel is based on traditional loans. In both panels, each dot represents a zip code. In Panel A, dots are sized based on the number of loans in round 3 by the lender type in the given zip code, and the coloring is based on the ratio of flagged loans in round 3 by the lender type to the population in the given zip code. In Panel B, dots are sized based on the number of loans across all rounds by the lender type in the given zip code, and the coloring is based on the growth in lending from rounds 1 and 2 to round 3 by the lender type in the given zip code. In both panels, zip codes are filtered to those with at least 25 loans. The dashed lines are linear fits and correlations are shown at the bottom left corner of each subpanel.

Panel A. Flagged Loans in Round 3



Panel B. Lending Growth

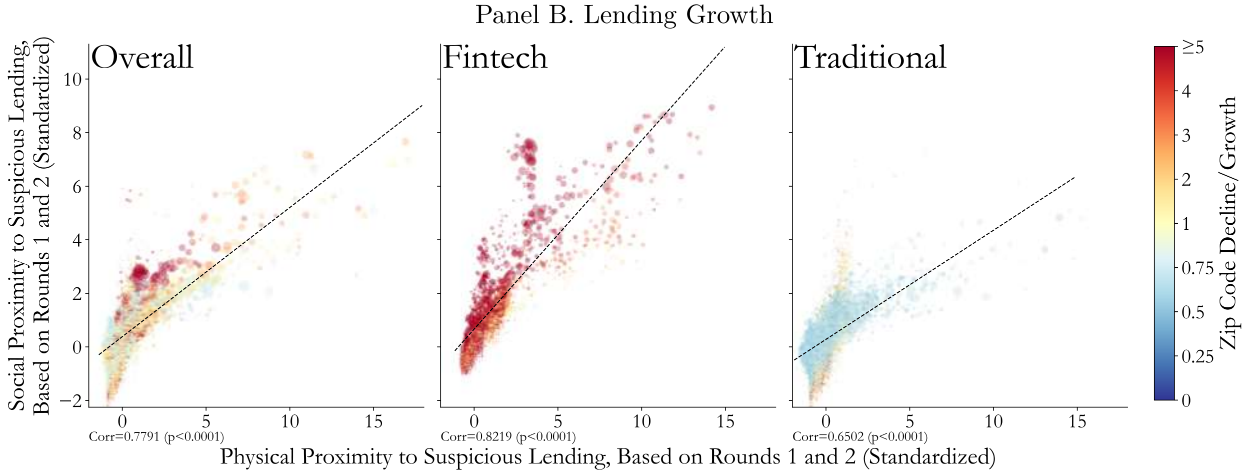
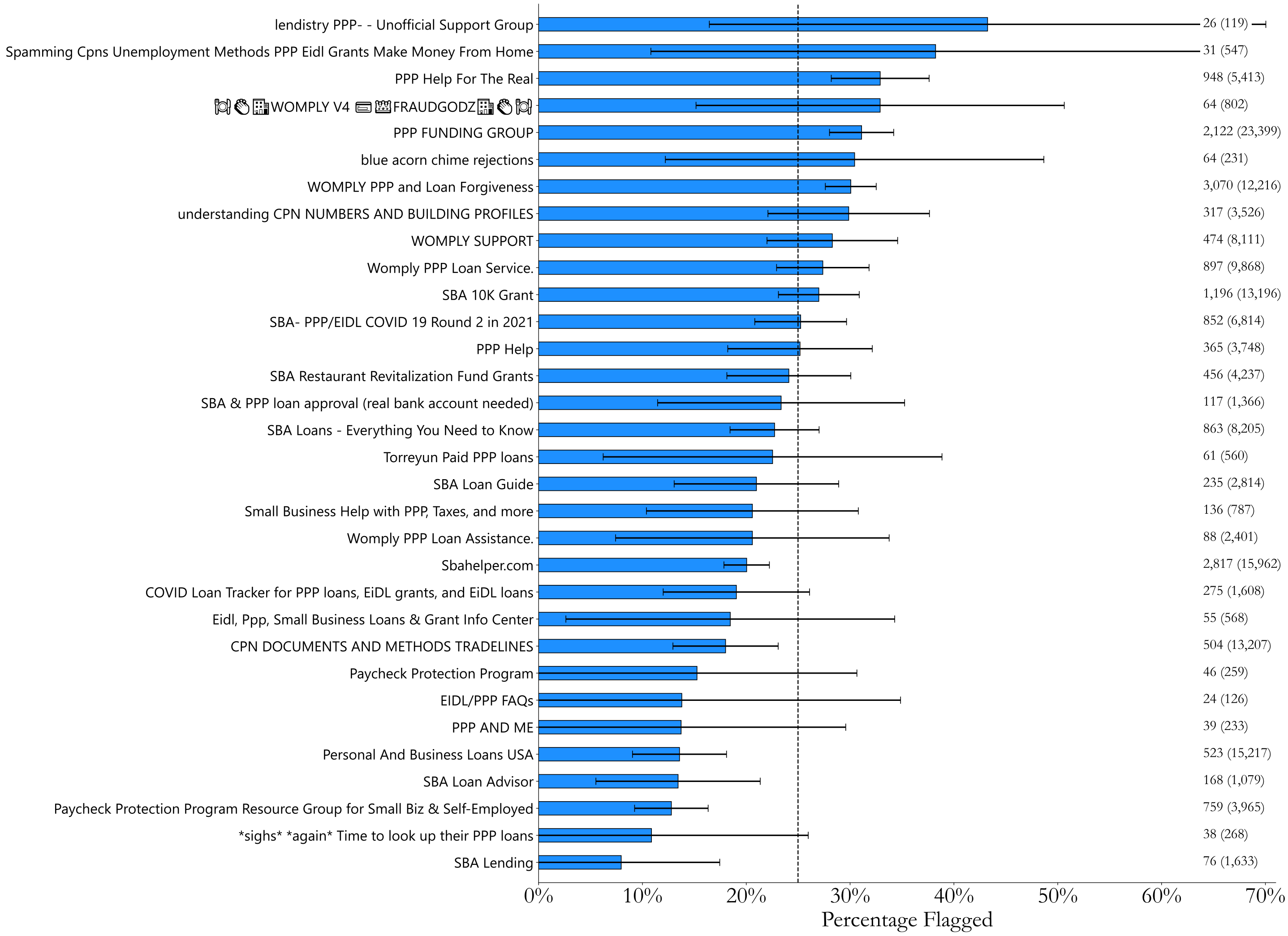


Figure IA.10. Social Media Groups, Alternative Search

This figure replicates Figure 7 using groups found by searching the social media platform for the following terms: "PPP," "Paycheck Protection Program," "SBA," and "Small Business Administration."

Panel A. Percentage Flagged by Group



Panel B. Wordcloud of Bottom (Left) and Top (Right) Ten Groups

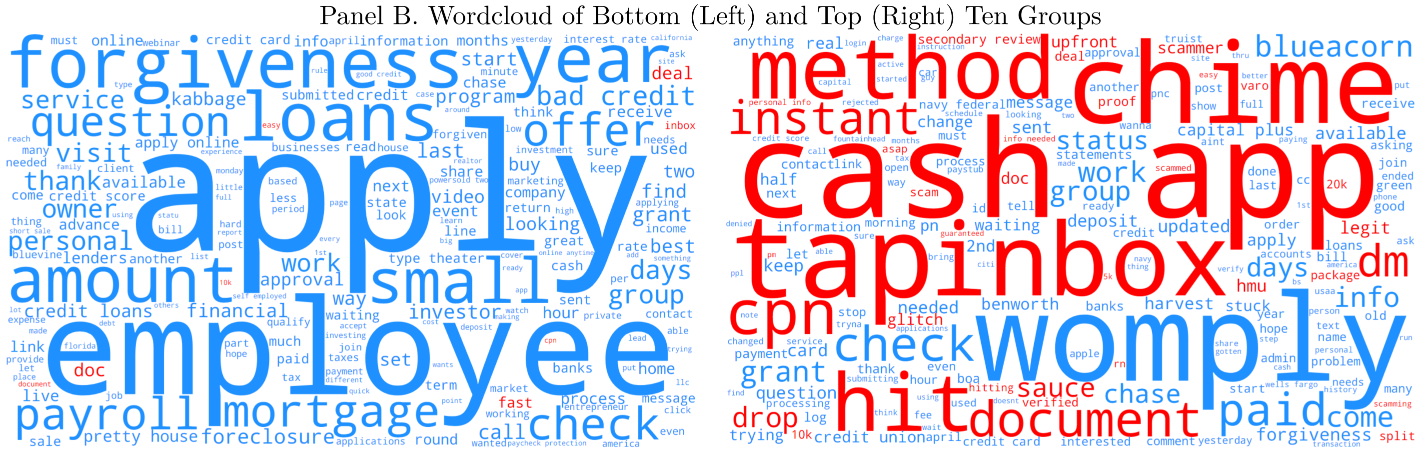


Figure IA.11. Social Media Groups, All Groups

This figure shows the complete version of Figure 7, Panel A.

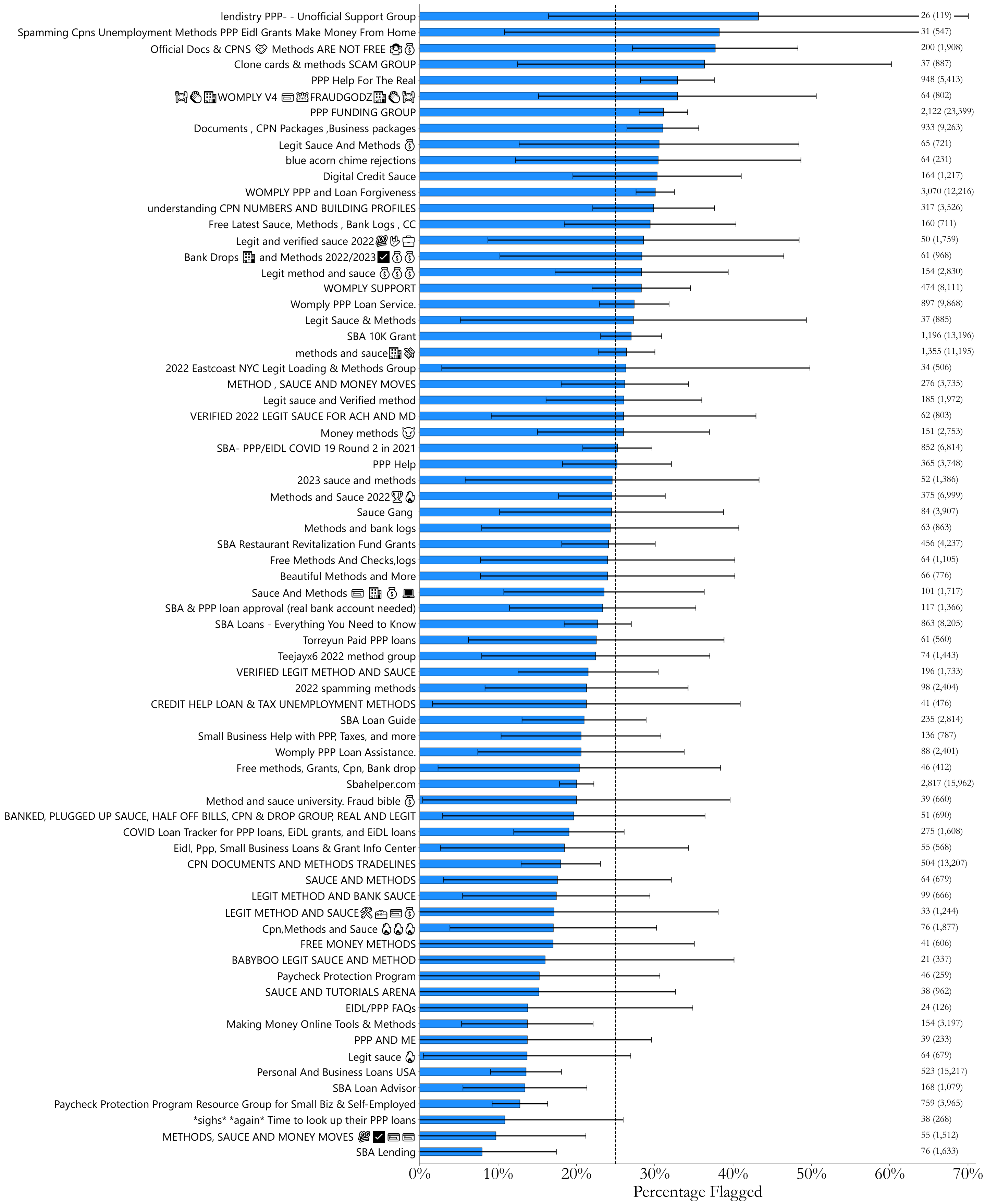


Figure IA.12. Social Media Groups, Group Fixed Effects

This figure shows the effect of individuals being in specific social media groups on the likelihood that their loan is flagged by at least one primary flag. Fixed effects for each group are shown from a regression of an indicator for whether the loan is flagged on indicators for whether the borrower is in each group. The green bars are from a regression without any additional fixed effects or controls. Blue bars are from a regression that controls for CBSA. Orange bars are from a regression that controls for log jobs reported, log loan amount, CBSA, and business type. The number of members in each group that are matched to a PPP loan and, in parentheses, the number of members in each group are noted to the right of each set of bars. The F-statistics, and corresponding p-values, for the joint significance of each set of fixed effects are shown below the legend.

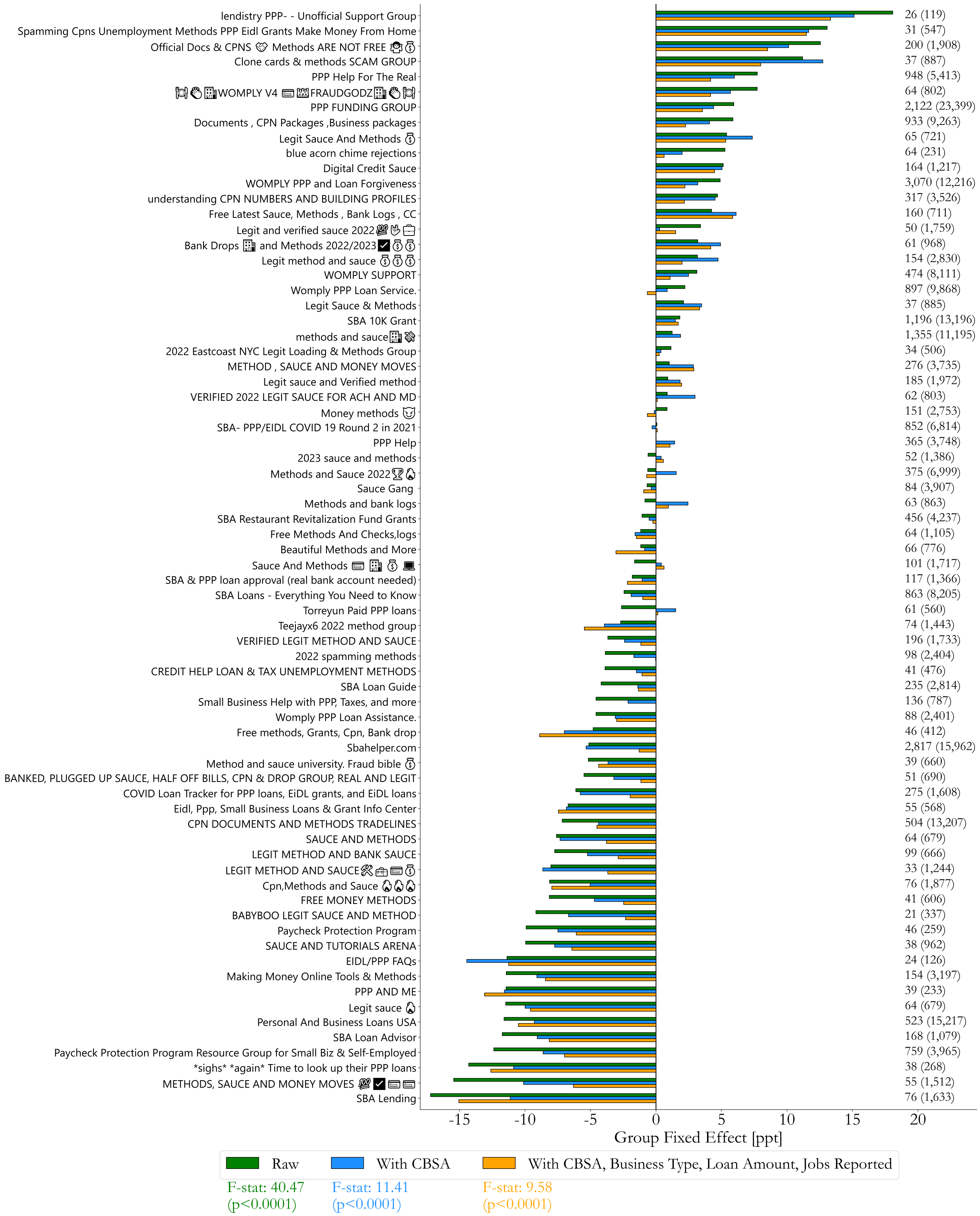


Figure IA.13. Social Proximity to Womply and BlueAcorn

This figure shows the relationship between social proximity to Womply and BlueAcorn loans and their relation with Womply and BlueAcorn loans per capita across zip codes. Loans originated by Capital Plus are excluded from the calculation of social proximity and loans per capita for both Womply and BlueAcorn. Both subpanels plot social proximity to Womply loans versus social proximity to BlueAcorn loans. Each dot is a zip code and is sized based on the total number of loans originated by either Womply or BlueAcorn in the zip code. In the left (right) subpanel, coloring is based on the number of Womply (BlueAcorn) loans per capita. Zip codes are filtered to those with at least 25 PPP loans.

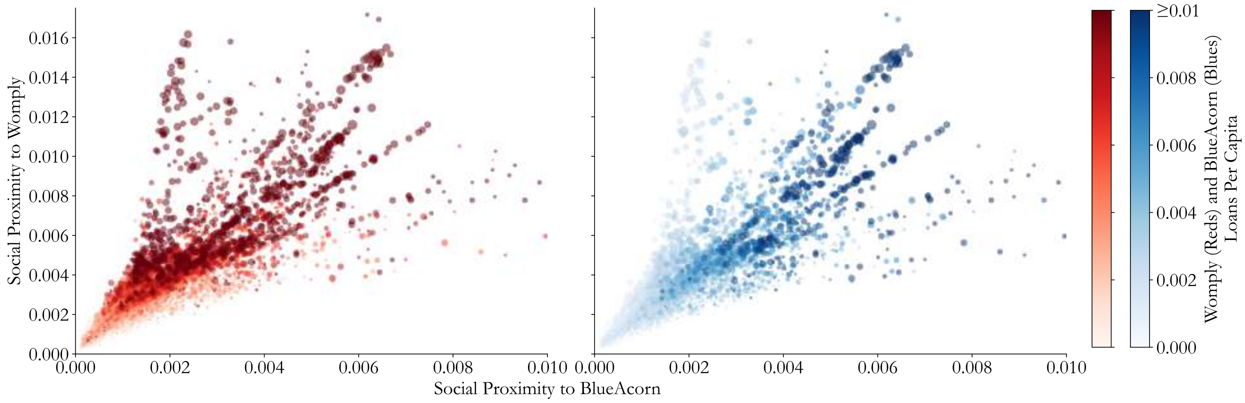


Figure IA.14. EIDL Agents, Within Zip Code Simulations

This figure replicates Figure 9, Panel A with the simulations rerun assuming each loan has an independent probability equal to the fraud rate in its zip code of being flagged as suspicious. Agents with at least 25 loans are considered.

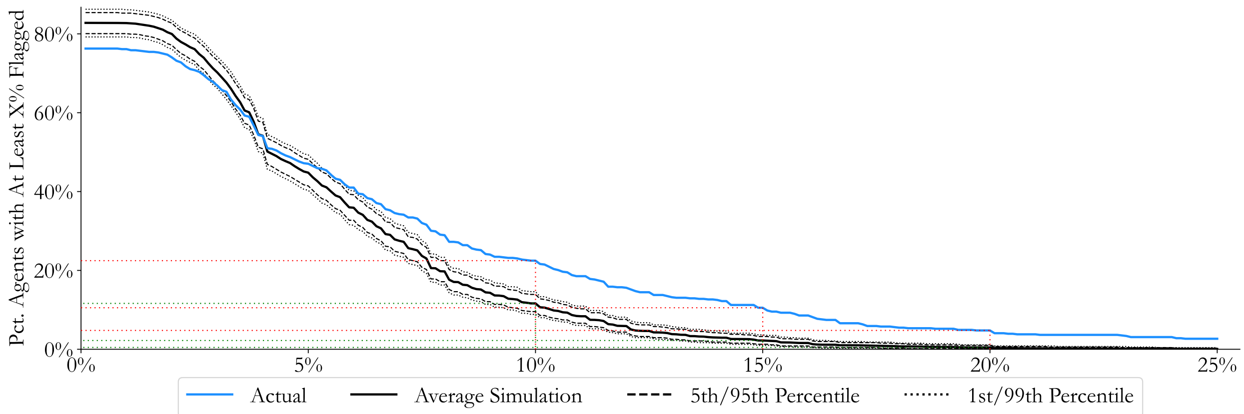


Figure IA.15. Google Trends, Interest in ERC

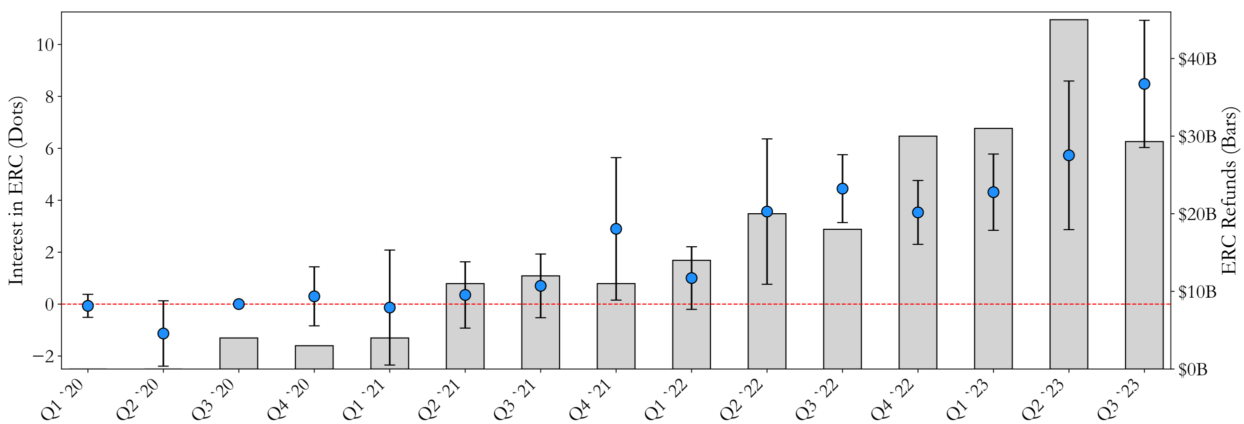
This figure shows additional details regarding the relationship between PPP fraud and Google searches related to the ERC across geographic areas. Panel A examines the connection between PPP fraud and search activity related to the ERC during each quarter from Quarter 1 of 2020 to Quarter 3 of 2023, as well as the amount of ERC refunds provided each quarter based on Wall Street Journal (2023a). The following regression is estimated and β_q is plotted:

$$Interest\ in\ ERC_{q,d} = \sum_{q \neq Q3\ 2020} \beta_q (Flagged\ Per\ Capita_d \times 1(Quarter = q)) + Quarter_q + DMA_d$$

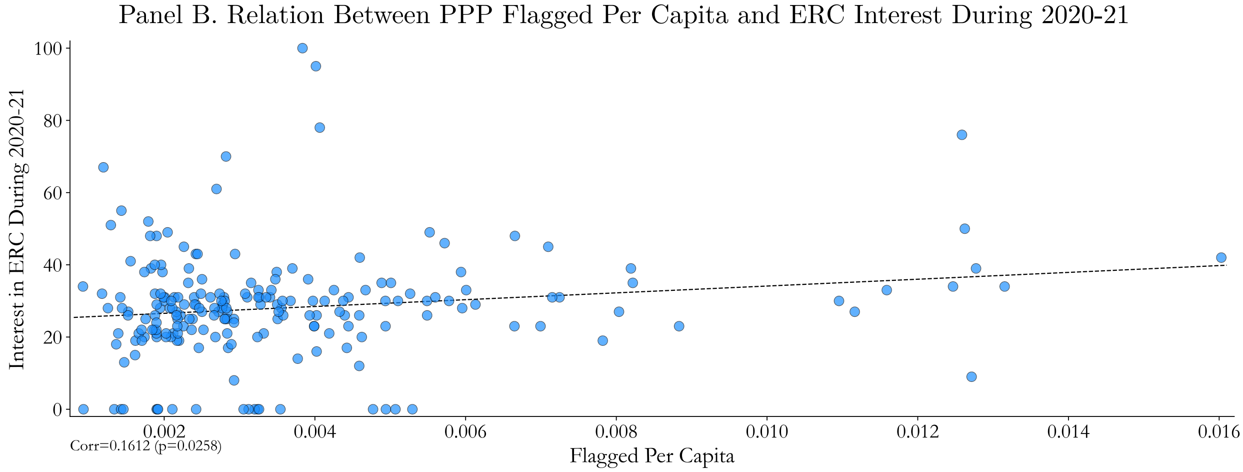
where q is a quarter and d is a Designated Marketing Area (DMA), $Interest\ in\ ERC_{q,d}$ is the Google Trends interest in ERC in DMA d during quarter q, $Flagged\ Per\ Capita_d$ is the number of flagged PPP loans per capita in DMA d, $Quarter_q$ is quarter fixed effect, and DMA_d is a DMA fixed effect. $Flagged\ Per\ Capita_d$ is standardized to have a mean of zero and a standard deviation of 1. The errorbars represent 95% confidence intervals based on standard errors that are double clustered by DMA and quarter.

Panel B examines the connection between PPP fraud and search activity for the ERC during 2020-21. Panel C replicates Figure 10, Panel B and Panel B of this figure controlling for 2018-19 searches related to the ERC (both subpanels) and 2020-21 searches related to the ERC (right subpanel). Panel D examines the relationship between Google searches for ERC in 2022-23 and search activity related to the PPP and the two largest FinTech PPP platforms (Womply and BlueAcorn).

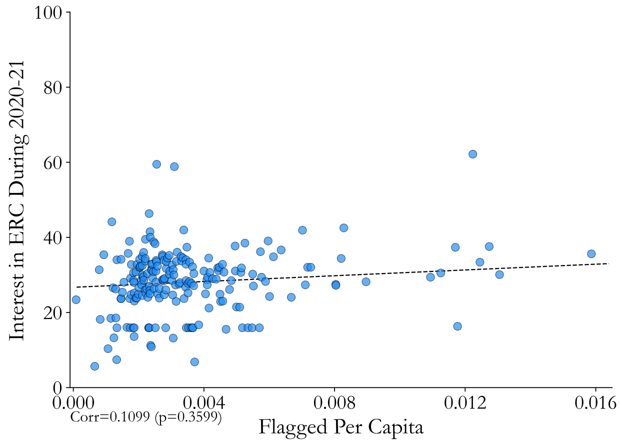
Panel A. Relation Between PPP Flagged Per Capita and ERC Interest Over Time

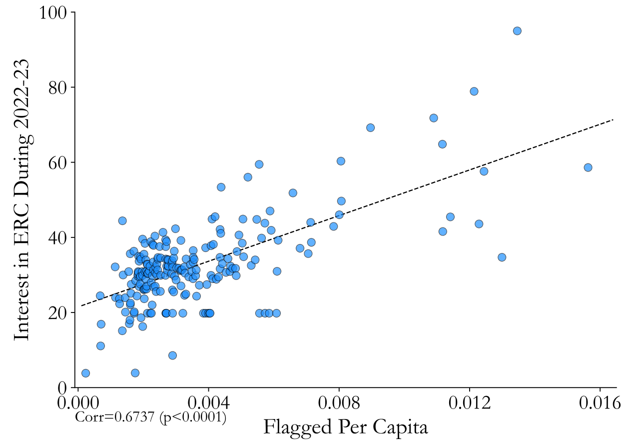


Panel B. Relation Between PPP Flagged Per Capita and ERC Interest During 2020-21

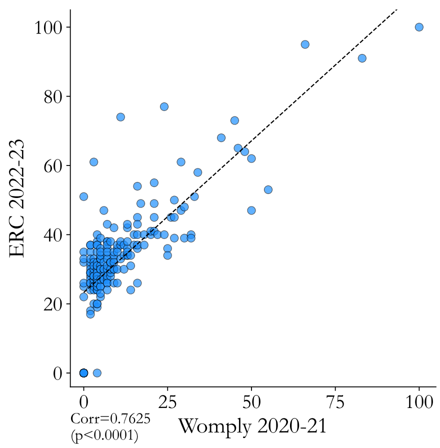


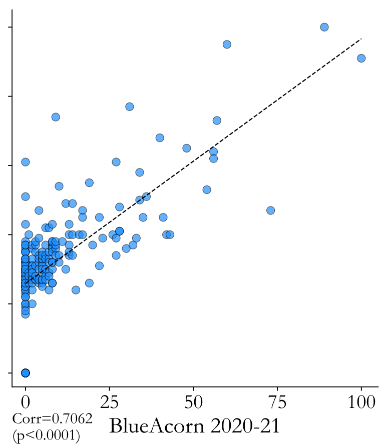
Panel C. Relation Between PPP Flagged Per Capita and ERC Interest During 2020-21 and 2022-23, Controlling for Past ERC Interest





Panel D. Correlation Between Google Trends Interest in PPP and ERC





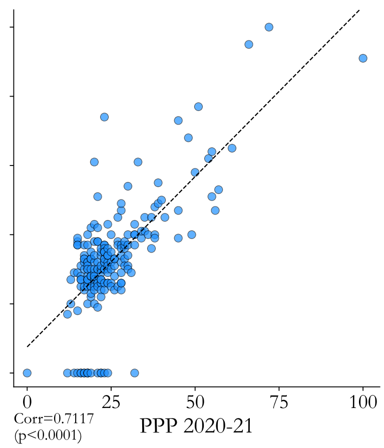


Table IA.I. EIDL and EIDL Advance Flags, Odds Ratios

This table examines odds ratios between different indicators of suspicious lending in the Advance component (Panel A) and loan component (Panel B) of the EIDL. The business registry, multiple loans, and EIDL > PPP jobs flags are as defined in Griffin, Kruger, and Mahajan (2023). *Advance without EIDL* flag is an indicator variable for whether the EIDL Advance does not have a corresponding EIDL loan. Note that odds ratios are symmetric, which is why only values for the lower triangular are provided. Robust standard errors are clustered by zip code.

Panel A. EIDL Advance

	Business Registry	Multiple Loans	EIDL > PPP Jobs	Advance Without EIDL
Business Registry	-			
Multiple Loans	1.157*** (4.14)			
EIDL > PPP Jobs	1.393*** (8.26)	2.661*** (23.73)	-	
Advance Without EIDL	1.377*** (28.75)	1.107*** (7.75)	2.168*** (28.76)	-

Panel B. EIDL

	Business Registry	Multiple Loans	EIDL > PPP Jobs
Business Registry	-		
Multiple Loans	1.276*** (6.68)	-	
EIDL > PPP Jobs	1.172*** (4.65)	1.533*** (11.17)	-

z-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table IA.II. Relationship Between Zip Code Level Fraud Rates in PPP and EIDL/EIDL Advance

This table examines the zip code-level relationship between fraud rates in the PPP and the Advance component (Panel A) or loan component (Panel B) of the EIDL. The dependent variables used are shown at the top of each column and are measures of suspicious lending in the Advance component (Panel A) or loan component (Panel B) of the EIDL program. The independent variable used in both panels is the percentage of PPP loans flagged by at least one of the business registry, multiple loans, or high compensation flags. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the number of EIDL advances (EIDL loans) for Panel A (Panel B) in each zip code. All variables (both independent and dependent) are standardized at the zip code level to have a mean of 0 and a standard deviation of 1. Zip codes are filtered to those with at least 25 PPP loans and 25 EIDL Advances (EIDL loans) for Panel A (Panel B). Fixed effects and controls are as indicated at the bottom of each column. Robust standard errors are clustered by county.

Panel A. EIDL Advance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dep. Variable:	Pct. Advances Without EIDL			Pct. Advances With Jobs > PPP Jobs			Pct. Advances Flagged		
Pct. PPP Flagged	0.661***	0.711***	0.497***	0.804***	0.893***	0.768***	0.352***	0.349***	0.449***
	(9.41)	(20.27)	(6.73)	(5.49)	(8.91)	(7.66)	(8.35)	(4.95)	(4.28)
County FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Demographics	No	No	Yes	No	No	Yes	No	No	Yes
Observations	16,036	16,036	16,036	16,036	16,036	16,036	16,036	16,036	16,036
Num. Counties	2,032	2,032	2,032	2,032	2,032	2032	2,032	2,032	2032
R^2	0.437	0.680	0.715	0.646	0.857	0.867	0.124	0.409	0.477

Panel B. EIDL

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Variable:	Pct. EIDL With Jobs > PPP Jobs			Pct. EIDL Flagged		
Pct. PPP Flagged	0.558***	0.619***	0.533***	0.268***	0.0832**	0.226***
	(7.11)	(7.21)	(6.77)	(4.14)	(2.45)	(3.95)
County FE	No	Yes	Yes	No	Yes	Yes
Demographics	No	No	Yes	No	No	Yes
Observations	13,248	13,248	13,248	13,248	13,248	13,248
Num. Counties	1,596	1,596	1,596	1,596	1,596	1,596
R^2	0.312	0.497	0.503	0.0719	0.473	0.524

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table IA.III. Unemployment Insurance Claims

This table examines the relationship between suspicious PPP lending rates and initial unemployment insurance claims across counties. The demographic variables used as controls are the percentage of adults with at least a bachelor's degree, median income, pre-pandemic unemployment rate, population density, percentage non-white, and poverty rate. *Pct. Each NAICS2* is the percentage of workers in the county working in each NAICS 2 digit classification. Panel B considers other economic variables. Columns (1) and (2) of Panel B consider changes in county-level GDP and income from 2019 to 2020 based on data from the Bureau of Economic Analysis. Columns (3), (4), and (5) consider the maximum decrease in employment, spending, and small business revenue based on data from the Economic Tracker by Opportunity Insights (described in Chetty et al. (2023)). The regressions are weighted by the number of individuals in the labor force as of December 2019. *Flagged Per Capita* is standardized to have a mean of 0 and a standard deviation of 1. Fixed effects and controls are as indicated at the bottom of each column. Robust standard errors are clustered by state.

Panel A. Initial Claims

Dep. Variable: Initial Claims in March to December 2020 Divided by Labor Force

	(1)	(2)	(3)	(4)
Flagged Per Capita	0.0471***	0.0431***	0.0582***	0.0338***
	(5.27)	(6.89)	(8.35)	(5.90)
Loans Per Capita		-0.00510	-0.0229***	-0.00944
		(-0.44)	(-5.08)	(-1.50)
State FE	Yes	Yes	Yes	Yes
State FE × Demographics	No	Yes	No	Yes
State FE × Pct. Each NAICS2	No	No	Yes	Yes
Observations	2,199	2,199	2,199	2,199
Num. States	33	33	33	33
R^2	0.853	0.946	0.951	0.972
Within State R^2	0.149	0.687	0.716	0.838
Mean Dep. Var.	0.424	0.424	0.424	0.424

Panel B. Placebo Tests

	(1)	(2)	(3)	(4)	(5)
	Change in		—— Max Decrease in ——		
Dep. Variable:	GDP	Income	Employment	Spending	Revenue
Flagged Per Capita	-0.00134	-0.00143	-0.00219	0.000853	0.0382
	(-0.36)	(-0.48)	(-0.30)	(0.05)	(1.62)
Loans Per Capita	-0.00183	0.000213	-0.00372	-0.0259**	-0.0553***
	(-0.42)	(0.06)	(-0.44)	(-2.29)	(-3.38)
State FE	Yes	Yes	Yes	Yes	Yes
State FE × Demographics	Yes	Yes	Yes	Yes	Yes
Observations	3,084	3,084	1,411	1,750	751
Num. States	50	50	50	50	50
R^2	0.442	0.705	0.467	0.553	0.746
Within State R^2	0.321	0.607	0.259	0.411	0.557
Mean Dep. Var.	-0.0331	0.0690	-0.232	-0.295	-0.452

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table IA.IV. Social Proximity to Suspicious Lending, Based on Percentage of Flagged Loans

This table replicates Table 2 based on percentage of flagged loans instead of flagged loans per capita. All proximity variables and the dependent variable are all based on percentage of flagged loans. The controls are the log of population density, percentage non-white, log of average household income, poverty rate, pre-pandemic unemployment, percentage with college education, and the shares of the friends of Facebook users in the zip code who live within 50 and 150 miles of them. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the number of PPP loans in each zip code. All variables (both independent and dependent) are standardized at the zip code level to have a mean of 0 and a standard deviation of 1. Zip codes are filtered to those with at least 25 loans. Fixed effects and controls are as indicated at the bottom of each column. Robust standard errors are clustered by county.

Panel A. Social and Physical Proximity

Dep. Variable: Percentage of Loans Flagged By At Least One Primary Measure in Zip Code

	(1)	(2)	(3)	(4)
Social Proximity to Suspicious Lending	1.215***	1.011***		1.068***
	(41.63)	(35.43)		(33.70)
Physical Proximity to Suspicious Lending			0.890***	-0.137**
			(5.62)	(-2.23)
County FE	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	Yes
Observations	19,753	19,753	19,753	19,753
Num. Counties	2,338	2,338	2,338	2,338
R^2	0.881	0.889	0.837	0.890
Within County R^2	0.696	0.717	0.582	0.718

Panel B. Homophily and Instrumental Variables

Dep. Variable: Percentage of Loans Flagged By At Least One Primary Measure in Zip Code

	(1)	(2)	(3)	(4)
Method:	OLS		IV	
Instrument:			$\geq$ 100 Mi	$\geq$ 500 Mi
Social Proximity to Suspicious Lending		1.157***	1.137***	1.101***
		(39.10)	(17.03)	(14.83)
Demographic Proximity to Suspicious Lending	0.479***	0.0553***		
	(5.92)	(2.98)		
County FE	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes
Observations	19,753	19,753	19,753	19,753
Num. Counties	2,338	2,338	2,338	2,338
R^2	0.729	0.882	0.888	0.888
Within County R^2	0.308	0.699	0.714	0.716
First Stage F-stat			41.08	30.59

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table IA.V. Social Proximity to Overall and Non-Suspicious Lending

This table replicates Table 2 for total and non-suspicious lending. The controls are the log of population density, percentage non-white, log of average household income, poverty rate, pre-pandemic unemployment, percentage with college education, the shares of the friends of Facebook users in the zip code who live within 50 and 150 miles of them, and the FinTech market share of PPP loans in the zip code. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the number of PPP loans in each zip code. All variables (both independent and dependent) are standardized at the zip code level to have a mean of 0 and a standard deviation of 1. Zip codes are filtered to those with at least 25 loans. Fixed effects and controls are as indicated at the bottom of each column. Robust standard errors are clustered by county.

Panel A. Overall Lending

Dep. Variable: Loans Per Capita in Zip Code

	(1)	(2)	(3)	(4)
Social Proximity to	0.384***	0.379***		0.300**
Overall Lending	(3.20)	(5.41)		(2.52)
Physical Proximity to			0.210**	0.104
Overall Lending			(2.29)	(0.89)
County FE	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	Yes
Observations	19,753	19,753	19,753	19,753
Num. Counties	2,338	2,338	2,338	2,338
R^2	0.212	0.360	0.355	0.362
Within County R^2	0.0233	0.207	0.200	0.209

Panel A. Non-Suspicious Lending

Dep. Variable: Non-Flagged Loans Per Capita in Zip Code

	(1)	(2)	(3)	(4)
Social Proximity to	0.300***	0.255**		0.134
Non-Suspicious Lending	(2.69)	(2.38)		(0.70)
Physical Proximity to			0.198**	0.150
Non-Suspicious Lending			(2.24)	(1.02)
County FE	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	Yes
Observations	19,753	19,753	19,753	19,753
Num. Counties	2,338	2,338	2,338	2,338
R^2	0.179	0.340	0.342	0.344
Within County R^2	0.0130	0.206	0.209	0.210

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table IA.VI. Social and Physical Proximity to Suspicious Lending, Zip Code Level

This table examines the relationship between social/physical proximity to suspicious lending and flagged per capita at the zip code level. The controls are the log of population density, percentage non-white, log of average household income, poverty rate, pre-pandemic unemployment, percentage with college education, the shares of the friends of Facebook users in the zip code who live within 50 and 150 miles of them, and the FinTech market share of PPP loans in the zip code. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the number of loans in each zip code. All variables (both independent and dependent) are standardized at the zip code level to have a mean of 0 and a standard deviation of 1. Zip codes are filtered to those with at least 25 loans. Fixed effects and controls are as indicated at the bottom of each column. Robust standard errors are clustered by county.

Panel A. Without Controls

Dep. Variable: Flagged By At Least One Primary Measure Per Capita

	(1)	(2)	(3)	(4)	(5)	(6)
Social Proximity	0.838***		0.924***	1.110***		1.137***
to Susp. Lending	(16.53)		(9.78)	(28.62)		(18.83)
Physical Proximity		0.637***	-0.106		0.887*	-0.0779
to Susp. Lending		(3.98)	(-1.27)		(1.94)	(-1.01)
County FE	No	No	No	Yes	Yes	Yes
Observations	19,753	19,753	19,753	19,753	19,753	19,753
Num. Counties	2,338	2,338	2,338	2,338	2,338	2,338
R^2	0.702	0.406	0.706	0.773	0.563	0.774
Within County R^2				0.555	0.143	0.556

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Panel B. With Controls

Dep. Variable: Flagged By At Least One Primary Measure Per Capita

	(1)	(2)	(3)	(4)	(5)	(6)
Social Proximity to Susp. Lending	0.859***		0.916***	1.071***		1.116***
	(20.54)		(11.00)	(47.46)		(33.66)
Physical Proximity to Susp. Lending		0.520***	-0.0605		0.646**	-0.0959
		(5.13)	(-0.92)		(2.57)	(-1.33)
Log Population Density	-0.0565**	-0.168***	-0.0540**	-0.105***	-0.187***	-0.0968***
	(-2.56)	(-4.32)	(-2.45)	(-4.28)	(-3.74)	(-3.60)
Pct. Non-White	0.144***	0.315***	0.126***	0.187***	0.386***	0.179***
	(3.91)	(5.48)	(3.90)	(4.22)	(4.98)	(4.37)
Log Average Income	0.161***	0.144**	0.161***	0.201***	0.274***	0.200***
	(2.72)	(2.27)	(2.73)	(3.09)	(3.47)	(3.09)
Poverty Rate	0.0426	0.0179	0.0490*	0.0481*	0.0238	0.0528**
	(1.39)	(0.29)	(1.84)	(1.80)	(0.44)	(2.04)
Pre-Pandemic Unemployment	0.0831	0.242*	0.0726*	0.0513	0.206*	0.0482
	(1.59)	(1.76)	(1.71)	(1.01)	(1.73)	(0.98)
Pct. College Educated	-0.0834*	0.0911*	-0.0881**	-0.117**	-0.0176	-0.119**
	(-1.87)	(1.75)	(-1.98)	(-2.08)	(-0.25)	(-2.14)
Share Friends Within 50 mi	-0.0899***	-0.0382	-0.0827***	-0.158***	-0.258**	-0.152***
	(-2.78)	(-1.38)	(-2.80)	(-2.99)	(-2.06)	(-2.99)
Share Friends Within 150 mi	0.0139	0.0655*	-0.000408	0.0787	0.276	0.0665
	(0.54)	(1.68)	(-0.02)	(1.26)	(1.50)	(1.17)
Pct. FinTech	-0.144***	0.0741**	-0.139***	-0.151***	0.179***	-0.161***
	(-3.60)	(2.14)	(-4.01)	(-4.11)	(2.71)	(-4.13)
County FE	No	No	No	Yes	Yes	Yes
Observations	19,753	19,753	19,753	19,753	19,753	19,753
Num. Counties	2,338	2,338	2,338	2,338	2,338	2,338
R^2	0.735	0.583	0.736	0.796	0.693	0.797
Within County R^2				0.600	0.398	0.601

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table IA.VII. Social and Physical Proximity to Suspicious Lending, Robustness to Detailed Racial and Ethnicity Controls

This table replicates Table IA.VI, Panel B with the percentage non-white control replaced with separate controls for percentage white, black, Asian, native, other race, and Hispanic/Latino. The other controls (log of population density, log of average household income, poverty rate, pre-pandemic unemployment, percentage with college education, the shares of the friends of Facebook users in the zip code who live within 50 and 150 miles of them, and the FinTech market share of PPP loans in the zip code) remain the same. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the number of loans in each zip code. All variables (both independent and dependent) are standardized at the zip code level to have a mean of 0 and a standard deviation of 1. Zip codes are filtered to those with at least 25 loans. Fixed effects and controls are as indicated at the bottom of each column. Robust standard errors are clustered by county.

Dep. Variable: Flagged By At Least One Primary Measure Per Capita

	(1)	(2)	(3)	(4)	(5)	(6)
Social Proximity	0.810***		0.868***	1.004***		1.043***
to Susp. Lending	(19.51)		(9.72)	(37.94)		(35.94)
Physical Proximity		0.473***	-0.0633		0.553**	-0.0849
to Susp. Lending		(5.66)	(-0.93)		(2.44)	(-1.38)
Pct. White	-0.0257	-0.0541	0.0366	0.101	0.466	0.116
	(-0.22)	(-0.40)	(0.25)	(0.49)	(1.40)	(0.55)
Pct. Black	0.184	0.341**	0.231	0.333	0.945**	0.342
	(1.37)	(2.07)	(1.46)	(1.52)	(2.50)	(1.54)
Pct. Asian	0.0179	0.0187	0.0377	0.0776	0.225*	0.0813
	(0.32)	(0.32)	(0.59)	(1.00)	(1.83)	(1.03)
Pct. Native	-0.00480	0.00134	0.00164	0.00391	0.0244	0.00592
	(-0.33)	(0.07)	(0.10)	(0.18)	(0.77)	(0.27)
Pct. Other Race	-0.0161	0.0270	-0.00991	0.0160	0.123	0.0179
	(-0.46)	(0.53)	(-0.29)	(0.25)	(1.23)	(0.28)
Pct. Hispanic/	-0.0269	-0.0985***	-0.0141	-0.0431	-0.00982	-0.0396
Latino	(-1.22)	(-2.59)	(-0.71)	(-1.03)	(-0.16)	(-0.98)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
County FE	No	No	No	Yes	Yes	Yes
Observations	19,753	19,753	19,753	19,753	19,753	19,753
Num. Counties	2,338	2,338	2,338	2,338	2,338	2,338
R^2	0.744	0.616	0.745	0.803	0.721	0.803
Within County R^2				0.613	0.452	0.614

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table IA.VIII. Social and Physical Proximity to Suspicious Lending, Loan Level

This table examines the relationship between social/physical proximity to suspicious lending and whether a loan is flagged by at least one primary flag. In columns (1)-(3), social and physical proximity to suspicious lending is calculated based on all loans; in columns (4) and (5), they are calculated based on only FinTech and traditional loans, respectively. Proximity measures are standardized to have a mean of 0 and a standard deviation of 1. Zip codes with at least 25 loans are used in the calculation of social and physical proximity to suspicious lending. The loan sample used in each column is denoted at the top of each column. The demographic controls are log population density, percentage non-white, log of average household income, poverty rate, pre-pandemic unemployment, and percentage with college education in the zip code. *Pct. Friends 50 & 150 Mi* are the share of the friends of Facebook users in the zip code who live within 50 and 150 miles of them. Fixed effects and controls are as indicated at the bottom of each column. Robust standard errors are double clustered by zip code and lender.

Dep. Variable: 1(Flagged by At Least One Primary Flag)

	(1)	(2)	(3)	(4)	(5)
Loan Sample:		All		FinTech	Traditional
Social Proximity to Suspicious Lending	0.0257***		0.0282***	0.0204***	0.000716
	(9.51)		(9.04)	(5.70)	(0.68)
Physical Proximity to Suspicious Lending		0.0126***	-0.00569***	-0.00395	-0.000343
		(4.96)	(-3.85)	(-1.59)	(-0.88)
Demographic Controls	Yes	Yes	Yes	Yes	Yes
Pct. Friend 50 & 150 Mi	Yes	Yes	Yes	Yes	Yes
ln(Jobs Reported)	Yes	Yes	Yes	Yes	Yes
ln(Loan Amount)	Yes	Yes	Yes	Yes	Yes
Lender FE	Yes	Yes	Yes	Yes	Yes
Business Type FE	Yes	Yes	Yes	Yes	Yes
NAICS × CBSA FE	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes
Observations	10,517,492	10,517,492	10,517,492	3,627,469	6,859,330
Num. Lenders	4,743	4,743	4,743	147	4,595
R^2	0.299	0.298	0.299	0.427	0.120
Mean of Dep. Var.	0.130	0.130	0.130	0.234	0.0748

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table IA.IX. Excess Share at County-Level

This table shows the relationship between excess share and proximity to excess share at the county level. The demographic controls are log of population density, percentage non-white, log of average household income, poverty rate, pre-pandemic unemployment, and percentage with college education in the county. *Pct. Friends 50 & 150 Mi* are the share of the friends of Facebook users in the county who live within 50 and 150 miles of them. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the number of loans in each county. All variables (both independent and dependent) are standardized at the county level to have a mean of 0 and a standard deviation of 1. Fixed effects and control variables are as indicated at the bottom of each column. Robust standard errors are clustered at the state level.

Dep. Variable: Excess Loans Per Capita

	(1)	(2)	(3)	(4)
Social Proximity	0.747***	0.468***		0.519***
to Excess Share	(7.42)	(5.74)		(3.71)
Physical Proximity			0.432	-0.135
to Excess Share			(1.13)	(-0.27)
State FE	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	Yes
Observations	3,187	3,187	3,187	3,187
R^2	0.597	0.691	0.667	0.691
Within State R^2	0.220	0.402	0.355	0.403

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table IA.X. Social and Physical Proximity to Suspicious Lending, Individual Flag

This table replicates column (3) from Table IA.VIII for each of the primary flags individually. Proximity measures are standardized to have a mean of 0 and a standard deviation of 1. Fixed effects and control variables are as indicated at the bottom of each column. Robust standard errors are double clustered by zip code and lender.

	(1)	(2)	(3)	(4)
Flag:	Business Registry	Multiple Loans	High Comp.	EIDL > PPP Jobs
Social Proximity to Flag	0.00474***	0.00797***	0.0143***	0.0411***
	(5.23)	(5.69)	(3.53)	(9.16)
Physical Proximity to Flag	-0.00136***	0.00265	0.00486**	-0.00373**
	(-2.80)	(1.23)	(2.41)	(-2.26)
Demographic Controls	Yes	Yes	Yes	Yes
Pct. Friend 50 & 150 Mi	Yes	Yes	Yes	Yes
ln(Jobs Reported)	Yes	Yes	Yes	Yes
ln(Loan Amount)	Yes	Yes	Yes	Yes
Lender FE	Yes	Yes	Yes	Yes
Business Type FE	Yes	Yes	Yes	Yes
NAICS × CBSA FE	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes
Observations	5,207,885	10,517,492	3,270,105	2,586,514
Num. Lender	4,542	4,743	4,551	4,481
R^2	0.104	0.0393	0.592	0.282
Mean of Dep. Var.	0.0506	0.0172	0.238	0.0900

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table IA.XI. Social and Physical Proximity to Suspicious Lending, County Level

This table replicates Table IA.VIII using social and physical proximity to suspicious lending calculated based on social connections between counties. In columns (1)-(3), social and physical proximity to suspicious lending are calculated based on all loans; in columns (4) and (5), they are calculated based on only FinTech and traditional loans, respectively. Proximity measures are standardized to have a mean of 0 and a standard deviation of 1. The loan sample used in each column is denoted at the top of each column. The demographic controls are log population density, percentage non-white, log of average household income, poverty rate, pre-pandemic unemployment, and percentage with college education in the zip code. *Pct. Friends 50 & 150 Mi* are the share of the friends of Facebook users in the zip code who live within 50 and 150 miles of them. Fixed effects and control variables are as indicated at the bottom of each column. Robust standard errors are double clustered by county and lender.

Dep. Variable: 1(Flagged by At Least One Primary Flag)

	(1)	(2)	(3)	(4)	(5)
Loan Sample:		All		FinTech	Traditional
Social Proximity to Suspicious Lending	0.0130***		0.0151***	0.0230***	0.00193*
	(5.94)		(6.09)	(5.18)	(1.90)
Physical Proximity to Suspicious Lending		0.0114***	-0.00529	-0.00903	-0.00593**
		(2.93)	(-1.26)	(-1.01)	(-2.52)
Demographic Controls	Yes	Yes	Yes	Yes	Yes
Pct. Friend 50 & 150 Mi	Yes	Yes	Yes	Yes	Yes
ln(Jobs Reported)	Yes	Yes	Yes	Yes	Yes
ln(Loan Amount)	Yes	Yes	Yes	Yes	Yes
Lender FE	Yes	Yes	Yes	Yes	Yes
Business Type FE	Yes	Yes	Yes	Yes	Yes
NAICS × State FE	Yes	Yes	Yes	Yes	Yes
Observations	11,386,662	11,386,662	11,386,662	3,729,480	7,655,829
Num. Lenders	4,777	4,777	4,777	147	4,630
R^2	0.249	0.249	0.249	0.369	0.0814
Mean of Dep. Var.	0.124	0.124	0.124	0.230	0.0719

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table IA.XII. Social and Physical Proximity to Suspicious Lending, Based on Chicago Taxi & Rideshare Trips

This table examines the relationship between social and physical proximity to suspicious lending based on an alternative measure and whether a loan is flagged in Chicago. Specifically, we use taxi trip data from 2013 to 2019 and rideshare trip data from 2018 to 2019 (see here and here). Analogous to the Social Connectedness Index from Bailey et al. (2018a, 2020), we calculate the social connectedness between Census Tracts as $SCI^{Taxi}_{i,j} = \frac{Trips_{i,j}}{\sum_k Trips_{i,k} \times \sum_k Trips_{k,j}}$ where $Trips_{i,j}$ is the number of taxi and rideshare trips between Census Tract i and j. Then, $SCI^{Taxi}_{i,j}$ is used to calculate social proximity to suspicious lending in the same way as done previously using the Social Connectedness Index from Bailey et al. (2018a, 2020). Proximity measures are standardized to have a mean of 0 and a standard deviation of 1. The demographic controls are population density, percentage non-white, median income, poverty rate, pre-pandemic unemployment, and percentage with college education in the Census Tract. Fixed effects and control variables are as indicated at the bottom of each column. Robust standard errors are double clustered by Census Tract and lender.

Dep. Variable: 1(Flagged by At Least One Primary Flag)

	(1)	(2)	(3)	(4)	(5)
Loan Sample:		All		FinTech	Traditional
Social Proximity to Susp. Lending	0.0345***		0.0310***	0.0292***	0.00249
	(6.33)		(4.17)	(3.81)	(0.76)
Physical Proximity to Susp. Lending		0.0319***	0.00590	0.00593	-0.00702
		(4.64)	(0.67)	(0.67)	(-1.44)
Demographic Controls	Yes	Yes	Yes	Yes	Yes
ln(Jobs Reported)	Yes	Yes	Yes	Yes	Yes
ln(Loan Amount)	Yes	Yes	Yes	Yes	Yes
Lender FE	Yes	Yes	Yes	Yes	Yes
Business Type FE	Yes	Yes	Yes	Yes	Yes
NAICS × Zip Code FE	Yes	Yes	Yes	Yes	Yes
Observations	153,707	153,707	153,707	110,797	41,081
Num. Lenders	508	508	508	49	451
R^2	0.543	0.542	0.543	0.509	0.321
Mean of Dep. Var.	0.357	0.357	0.357	0.469	0.0649

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table IA.XIII. Social and Physical Proximity to Suspicious Lending, EIDL Loans and Advance

This table replicates Table IA.VI for EIDL loans (Panel A) and Advances (Panels B and C) instead of PPP loans. The demographic controls are the log of population density, percentage non-white, log of average household income, poverty rate, pre-pandemic unemployment, and percentage of adults with a college education in the zip code. *Pct. Friends 50 & 150 Mi* are the share of the friends of Facebook users in the zip code who live within 50 and 150 miles of them. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the number of EIDL loans (Advances) for Panel A (Panels B and C) in each zip code. All variables (both independent and dependent) are standardized to have a mean of 0 and a standard deviation of 1. Zip codes are filtered to those with at least 25 EIDL loans (Advance) for Panel A (Panels B and C). Fixed effects and control variables are as indicated at the bottom of each column. Robust standard errors are clustered by county.

Panel A. EIDL Loans, Griffin, Kruger, and Mahajan (2023) Suspicious Lending Indicators

Dep. Variable: Flagged EIDL Loans Per Capita

	(1)	(2)	(3)	(4)	(5)	(6)
Social Proximity	0.605***		0.336**	0.677***		0.609***
to Flagged EIDL Loans	(7.96)		(2.27)	(6.44)		(5.53)
Physical Proximity		0.602***	0.316***		0.551***	0.142
to Flagged EIDL Loans		(18.45)	(2.75)		(4.24)	(1.45)
Demographic Controls	Yes	Yes	Yes	Yes	Yes	Yes
Pct. Friend 50 & 150 Mi	Yes	Yes	Yes	Yes	Yes	Yes
County FE	No	No	No	Yes	Yes	Yes
Observations	13,151	13,151	13,151	13,151	13,151	13,151
Num. Counties	1,576	1,576	1,576	1,576	1,576	1,576
R^2	0.420	0.417	0.442	0.510	0.484	0.511

Panel B. EIDL Advances, Griffin, Kruger, and Mahajan (2023) Suspicious Lending Indicators

Dep. Variable: Flagged EIDL Advances Per Capita

	(1)	(2)	(3)	(4)	(5)	(6)
Social Proximity	0.866***		0.973***	1.097***		1.207***
to Flagged EIDL Advances	(32.22)		(11.48)	(33.48)		(21.61)
Physical Proximity		0.665***	-0.117		1.144***	-0.253
to Flagged EIDL Advances		(9.28)	(-1.19)		(5.97)	(-1.52)
Demographic Controls	Yes	Yes	Yes	Yes	Yes	Yes
Pct. Friend 50 & 150 Mi	Yes	Yes	Yes	Yes	Yes	Yes
County FE	No	No	No	Yes	Yes	Yes
Observations	15,654	15,654	15,654	15,654	15,654	15,654
Num. Counties	1,929	1,929	1,929	1,929	1,929	1,929
R^2	0.774	0.606	0.776	0.820	0.704	0.822

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Panel C. EIDL Advances, Advances Without Loans

Dep. Variable: EIDL Advances Without EIDL Per Capita

	(1)	(2)	(3)	(4)	(5)	(6)
Social Proximity	0.657***		0.662***	0.805***		0.757***
to Excess Advances	(8.75)		(4.68)	(6.23)		(5.28)
Physical Proximity		0.540***	-0.00499		0.394***	0.0903
to Excess Advances		(7.87)	(-0.05)		(2.71)	(1.35)
Demographic Controls	Yes	Yes	Yes	Yes	Yes	Yes
Pct. Friend 50 & 150 Mi	Yes	Yes	Yes	Yes	Yes	Yes
County FE	No	No	No	Yes	Yes	Yes
Observations	15,654	15,654	15,654	15,654	15,654	15,654
Num. Counties	1,929	1,929	1,929	1,929	1,929	1,929
R^2	0.558	0.465	0.558	0.634	0.582	0.635

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table IA.XIV. Controlling for Demographic Proximity to Suspicious Lending

This table examines additional specifications of controlling for demographic proximity to suspicious lending, extending columns (1) and (2) of Table 2, Panel B. Columns (1) and (2) use demographic proximity to suspicious lending based on the average similarity between each pair of zip codes along six demographic variables. Columns (3) to (6) use proximity to suspicious lending based on the similarity between each pair of zip codes along six demographic variables separately. The similarity between zip code pairs for each demographic is defined as $1 - |PercentileRank(Demographic_i) - PercentileRank(Demographic_j)|$ where $PercetileRank(.)$ outputs a percentile rank between 0 and 1, i and j are two zip codes, and $|.|$ is the absolute value operator. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the number of PPP loans in each zip code. All variables (both independent and dependent) are standardized at the zip code level to have a mean of 0 and a standard deviation of 1. Zip codes are filtered to those with at least 25 loans. Fixed effects are as indicated at the bottom of each column. Robust standard errors are clustered by county.

Dep. Variable: Flagged By At Least One Primary Measure Per Capita

	(1)	(2)	(3)	(4)	(5)	(6)
Social Proximity to Suspicious Lending		0.832***		0.835***		1.104***
		(14.99)		(15.89)		(28.96)
Demographic Proximity to Suspicious Lending	0.430***	0.0118				
	(3.42)	(0.72)				
Median Income Proximity to Suspicious Lending			0.319***	0.00296	0.232**	0.0212
			(3.18)	(0.08)	(2.07)	(0.47)
Poverty Proximity to Suspicious Lending			-0.133	0.0722*	-0.0421	0.0660
			(-1.31)	(1.76)	(-0.48)	(1.62)
Pop. Density Proximity to Suspicious Lending			0.0186	-0.0948***	-0.183***	-0.135***
			(0.30)	(-3.59)	(-3.37)	(-6.09)
Pct. Non-White Proximity to Suspicious Lending			0.292***	0.0171	0.275***	0.0350
			(5.58)	(0.57)	(4.62)	(1.42)
Educ. Attainment Proximity to Suspicious Lending			0.153***	0.0939***	0.149***	0.0977***
			(4.15)	(5.57)	(6.65)	(6.51)
Unemployment Proximity to Suspicious Lending			0.169	0.0338	0.157	-0.00669
			(1.37)	(1.64)	(1.55)	(-0.35)
County FE	No	No	No	No	Yes	Yes
Observations	19,753	19,753	19,753	19,753	19,753	19,753
Num. Counties	2,338	2,338	2,338	2,338	2,338	2,338
R^2	0.185	0.702	0.211	0.711	0.574	0.778
Within County R^2					0.165	0.565

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table IA.XV. Relation Between Social and Demographic Proximity to Suspicious Lending

This table examines the relationships between social and demographic proximity to suspicious lending. Column use demographic proximity to suspicious lending based on the average similarity between each pair of zip codes along six demographic variables. Columns (2) to (8) use proximity to suspicious lending based on the similarity between each pair of zip codes along six demographic variables separately. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the number of PPP loans in each zip code. All variables (both independent and dependent) are standardized at the zip code level to have a mean of 0 and a standard deviation of 1. Zip codes are filtered to those with at least 25 loans. Fixed effects are as indicated at the bottom of each column. Robust standard errors are clustered by county.

Dep. Variable: Social Proximity to Suspicious Lending

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Demographic Proximity to Suspicious Lending	0.323***							
	(3.19)							
Median Income Proximity to Suspicious Lending		0.260***						0.191***
		(2.88)						(2.95)
Poverty Proximity to Suspicious Lending			0.218***					-0.0980*
			(2.98)					(-1.88)
Pop. Density Proximity to Suspicious Lending				0.122***				-0.0434
				(10.02)				(-0.91)
Pct. Non-White Proximity to Suspicious Lending					0.340***			0.217***
					(3.23)			(4.08)
Educ. Attainment Proximity to Suspicious Lending						-0.190**		0.0463**
						(-2.33)		(2.23)
Unemployment Proximity to Suspicious Lending							0.267**	0.148
							(2.42)	(1.63)
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	19,753	19,753	19,753	19,753	19,753	19,753	19,753	19,753
Num. Counties	2,338	2,338	2,338	2,338	2,338	2,338	2,338	2,338
R^2	0.820	0.810	0.801	0.774	0.815	0.792	0.816	0.833
Within County R^2	0.215	0.172	0.134	0.0143	0.192	0.0915	0.197	0.271

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table IA.XVI. Social Proximity to Suspicious Lending, Distant Zip Codes

This table examines additional specifications of instrumenting for social proximity based on all zip codes using distant zip codes (Panel A) and the reduced form of the IV regressions (Panel B), extending columns (3) and (4) of Table 2, Panel B. The version of social proximity used is denoted at the top of each column. In column (5) of Panel A, social connections to zip codes in non-overlapping rings (between 100 and 250 miles, between 250 and 500 miles, and over 500 miles) are used as instruments. The J-stat and p-value for an overidentification test are provided at the bottom of column (5) of Panel A. The controls are the log of population density, percentage non-white, log of average household income, poverty rate, pre-pandemic unemployment, percentage with college education, the shares of the friends of Facebook users in the zip code who live within 50 and 150 miles of them, and the FinTech market share of PPP loans in the zip code. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the number of PPP loans in each zip code. All variables (both independent and dependent) are standardized at the zip code level to have a mean of 0 and a standard deviation of 1. Zip codes and controls are filtered to those with at least 25 loans. Fixed effects and controls are as indicated at the bottom of each column. Robust standard errors are clustered by county.

Panel A. IV

Dep. Variable: Flagged By At Least One Primary Flag Per Capita

	(1)	(2)	(3)	(4)	(5)
Instrument:	$\geq$ 250 Mi	Outside County	Outside CBSA	Outside State	Concentric Rings
Social Proximity to Suspicious Lending	1.253***	1.080***	1.179***	1.111***	1.216***
	(15.48)	(15.86)	(19.78)	(17.82)	(14.94)
County FE	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Observations	19,753	19,753	17,498	19,753	19,717
Num. Counties	2,338	2,338	1,844	2,338	2,336
R^2	0.792	0.796	0.796	0.796	0.794
Within County R^2	0.592	0.600	0.602	0.600	0.595
First Stage F-stat	14.04	18.87	18.90	10.97	7.06
Hansen's J-stat					4.02
(p-value)					0.13

Panel B. Reduced Form

Dep. Variable: Flagged By At Least One Primary Flag Per Capita

	(1)	(2)	(3)	(4)	(5)	(6)
SP Based on:	$\geq$ 100 Mi	$\geq$ 250 Mi	$\geq$ 500 Mi	Outside County	Outside CBSA	Outside State
Social Proximity to Suspicious Lending	0.615***	0.581***	0.553***	0.982***	0.826***	0.787***
	(4.95)	(4.30)	(3.854)	(5.09)	(4.10)	(3.59)
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	19,753	19,753	19,753	19,753	17,498	19,753
Num. Counties	2,338	2,338	2,338	2,338	1,844	2,338
R^2	0.687	0.687	0.686	0.729	0.699	0.708
Within County R^2	0.386	0.385	0.385	0.469	0.414	0.427

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table IA.XVII. Social Proximity to Suspicious Lending, First Stage

This table shows the first stages of the IV regressions shown in columns (3) and (4) of Table 2, Panel B and all columns of Table IA.XVI, Panel A. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the number of PPP loans in each zip code. All variables (both independent and dependent) are standardized at the zip code level to have a mean of 0 and a standard deviation of 1. Zip codes and controls are filtered to those with at least 25 loans. Fixed effects and controls are as indicated at the bottom of each column. Robust standard errors are clustered by county.

Dep. Variable: Social Proximity to Suspicious Lending Based on All Zip Codes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
SP Based on Zip Codes $\geq$ 100 Mi	0.520***						
	(4.25)						
SP Based on Zip Codes $\geq$ 250 Mi		0.464***					
		(3.75)					
SP Based on Zip Codes $\geq$ 500 Mi			0.445***				0.378***
			(3.12)				(2.99)
SP Based on Zip Codes Outside County				0.910***			
				(4.35)			
SP Based on Zip Codes Outside CBSA					0.709***		
					(4.09)		
SP Based on Zip Codes Outside State						0.708***	
						(3.31)	
SP Based on Zip Codes $[100, 250)$ Miles							0.224***
							(3.67)
SP Based on Zip Codes $[250, 500)$ Miles							0.155***
							(2.79)
County FE	Yes	Yes	Yes	Yes	Yes	Yess	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yess	Yes
Observations	19,753	19,753	19,753	19,753	17,498	19,753	19,717
Num. Counties	2,338	2,338	2,338	2,338	1,844	2,338	2,336
R^2	0.566	0.555	0.556	0.742	0.609	0.652	0.588

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table IA.XVIII. Social Proximity to Suspicious Lending, Loan Level Based on Distant Zip Codes

This table examines the loan level IV and reduced form regressions that are equivalent to columns (3) and (4) of Table 2, Panel B. The version of social proximity used is denoted at the top of each column. In column (5) of Panel A, social connections to zip codes in non-overlapping rings (between 100 and 250 miles, between 250 and 500 miles, and over 500 miles) are used as instruments. The J-stat and p-value for an overidentification test are provided at the bottom of column (5) of Panel A. The demographic controls are the log of population density, percentage non-white, log of average household income, poverty rate, pre-pandemic unemployment, and percentage with college education. The independent variables are standardized at the zip code level to have a mean of 0 and a standard deviation of 1. Fixed effects and controls are as indicated at the bottom of each column. Robust standard errors are double clustered by zip code and lender.

Panel A. IV

Dep. Variable: 1(Flagged by At Least One Primary Flag)

Instrument:	(1) Outside CBSA	(2) $\geq$ 100 Mi	(3) $\geq$ 250 Mi	(4) $\geq$ 500 Mi	(5) Concentric Rings
Social Proximity to Suspicious Lending	0.0321*** (11.40)	0.0325*** (12.10)	0.0330*** (11.93)	0.0316*** (10.64)	0.0322*** (11.37)
Demographic Controls	Yes	Yes	Yes	Yes	Yes
Pct. Friend 50 & 150 Mi	Yes	Yes	Yes	Yes	Yes
ln(Jobs Reported)	Yes	Yes	Yes	Yes	Yes
ln(Loan Amount)	Yes	Yes	Yes	Yes	Yes
Lender FE	Yes	Yes	Yes	Yes	Yes
Business Type FE	Yes	Yes	Yes	Yes	Yes
NAICS × CBSA FE	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes
Observations	10,517,492	10,517,492	10,517,492	10,517,492	10,517,492
Num. Lenders	4,743	4,743	4,743	4,743	4,743
R^2	0.0681	0.0681	0.0680	0.0681	0.0683
Mean of Dep. Var.	0.130	0.130	0.130	0.130	0.130
First Stage F-stat	594.8	522.6	430.4	328.3	171.1
Hansen's J-stat					5.199
(p-value)					0.0743

Panel B. Reduced Form

Dep. Variable: 1(Flagged by At Least One Primary Flag)

SP Based on:	(1) Outside CBSA	(2) $\geq$ 100 Mi	(3) $\geq$ 250 Mi	(4) $\geq$ 500 Mi
Social Proximity to Suspicious Lending	0.0244*** (10.17)	0.0194*** (10.76)	0.0178*** (11.10)	0.0163*** (10.01)
(Same Fixed Effects and Controls as Panel A)				
Observations	10,517,492	10,517,492	10,517,492	10,517,492
Num. Lenders	4,743	4,743	4,743	4,743
R^2	0.298	0.298	0.298	0.298
Mean of Dep. Var.	0.130	0.130	0.130	0.130

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table IA.XIX. Social Capital, Without Social Proximity

This table examines the effect of social capital variables from Chetty et al. (2022a,b) on flagged per capita at the zip code level, i.e., it replicates Table 4 without including social proximity to suspicious lending. The demographic controls are the log of population density, percentage non-white, log of average household income, poverty rate, pre-pandemic unemployment, and percentage with college education in the zip code. *Pct. Friends 50 & 150 Mi* are the share of the friends of Facebook users in the zip code who live within 50 and 150 miles of them. *Pct. FinTech* is the FinTech market share of PPP loans in the zip code. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the number of PPP loans in each zip code. All variables (both independent and dependent) are standardized to have a mean of 0 and a standard deviation of 1. Zip codes are filtered to those with at least 25 loans. Fixed effects and controls are as indicated at the bottom of each column. Robust standard errors are clustered by county.

Dep. Variable: Flagged By At Least One Primary Flag Per Capita

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Economic Connectedness	-0.0157							0.0106
	(-0.36)							(0.19)
Exposure		-0.0170						
		(-0.50)						
Friending Bias			-0.0732					
			(-1.35)					
Clustering				-0.111				-0.104
				(-1.40)				(-1.20)
Support Ratio					-0.104***			-0.0759***
					(-3.62)			(-3.46)
Volunteering Rate						-0.0756**		-0.0984**
						(-2.49)		(-2.21)
Civic Organizations							0.152***	0.149***
							(7.06)	(6.59)
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demographic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pct. Friend 50 & 150 Mi	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pct. FinTech	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	16,814	16,814	16,814	16,814	16,814	16,814	16,814	16,814
Num. Counties	2,116	2,116	2,116	2,116	2,116	2,116	2,116	2,116
R^2	0.647	0.647	0.650	0.649	0.651	0.648	0.654	0.660

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table IA.XX. Monthly Spread Between Zip Codes and Counties

This table examines the spread of fraud between zip codes/counties over time. The unit of observation is zip code-month in Panel A and county-month in Panel B. The fraud rates are calculated based on loans originated during each month in the zip code/county. The demographic controls are population density, percentage non-white, median income, poverty rate, pre-pandemic unemployment, and percentage with college education in the zip code/county. All variables (both independent and dependent) are standardized to have a mean of 0 and a standard deviation of 1. Zip codes/counties with fewer than 100 loans during the entirety of the PPP and zip code/county-month observations with less than 10 loans during the given month are excluded. Fixed effects and control variables are as indicated at the bottom of each column. Robust standard errors are double clustered at zip code and month × county level in Panel A and double clustered at county and month × state level in Panel B.

Panel A. Between Zip Codes

Dep. Variable: Flagged By At Least One Primary Flag Per Capita

	(1)	(2)	(3)	(4)	(5)	(6)
Lagged Social Proximity to Suspicious Lending	0.604***		0.810***	0.363***		0.533***
	(4.82)		(5.09)	(3.24)		(3.68)
Lagged Physical Proximity to Suspicious Lending		0.342***	-0.482***		0.115**	-0.376***
		(3.76)	(-5.53)		(2.39)	(-4.93)
Lagged Flagged Per Capita				0.288***	0.360***	0.277***
				(8.22)	(6.86)	(8.35)
Zip Code FE	Yes	Yes	Yes	Yes	Yes	Yes
Month × County FE	Yes	Yes	Yes	Yes	Yes	Yes
Month × Pct. Fintech	Yes	Yes	Yes	Yes	Yes	Yes
Month × Demographics	Yes	Yes	Yes	Yes	Yes	Yes
Observations	92,086	92,086	92,086	92,086	92,086	92,086
R^2	0.831	0.809	0.834	0.848	0.841	0.851

Panel B. Between Counties

Dep. Variable: Flagged By At Least One Primary Flag Per Capita

	(1)	(2)	(3)	(4)	(5)	(6)
Lagged Social Proximity to Suspicious Lending	0.524***		0.610***	0.274***		0.354***
	(7.60)		(7.41)	(3.76)		(3.88)
Lagged Physical Proximity to Suspicious Lending		0.623**	-0.400**		0.169	-0.372**
		(2.89)	(-2.24)		(1.24)	(-2.24)
Lagged Flagged Per Capita				0.262***	0.296***	0.261***
				(4.47)	(5.29)	(4.46)
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Month × State FE	Yes	Yes	Yes	Yes	Yes	Yes
Month × Pct. Fintech	Yes	Yes	Yes	Yes	Yes	Yes
Month × Demographics	Yes	Yes	Yes	Yes	Yes	Yes
Observations	22,168	22,168	22,168	22,168	22,168	22,168
R^2	0.698	0.687	0.699	0.718	0.715	0.718

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table IA.XXI. Womply and BlueAcorn Usage and Mentions in Facebook Groups

This table examines whether members of social media groups are more likely to use a lender affiliated with Womply (columns (1) to (3)) or BlueAcorn (columns (4) to (6)) when the group has more discussion regarding Womply and BlueAcorn. *Womply Mentions Per Word* is defined as the number of times Womply or one of the lenders affiliated with it is mentioned in each group divided by the total number of words in the group. *BlueAcorn Mentions Per Word* is defined as the number of times BlueAcorn or one of the lenders affiliated with it is mentioned in each group divided by the total number of words in the group. Mentions of Capital Plus are not included in *Womply Mentions Per Word* nor *BlueAcorn Mentions Per Word* since Capital Plus is associated with both Womply and BlueAcorn. The independent variables are standardized to have a mean of 0 and a standard deviation of 1. Fixed effects and control variables are as indicated at the bottom of each column. Robust standard errors are double clustered at the social media group and individual level.

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Variable:	1(Womply)			1(BlueAcorn)		
Womply Mention Per Word	0.0229***		0.0318***		0.00267	-0.0103***
	(2.67)		(15.17)		(0.26)	(-6.95)
BlueAcorn Mentions Per Word		-0.000664	-0.0186***	0.0211***		0.0269***
		(-0.18)	(-15.66)	(6.25)		(8.52)
ln(Jobs Reported)	Yes	Yes	Yes	Yes	Yes	Yes
ln(Loan Amount)	Yes	Yes	Yes	Yes	Yes	Yes
Business Type FE	Yes	Yes	Yes	Yes	Yes	Yes
CBSA FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	26,713	26,713	26,713	26,713	26,713	26,713
R^2	0.131	0.121	0.135	0.127	0.117	0.129
Dep. Variable Mean	0.415	0.415	0.415	0.217	0.217	0.217

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table IA.XXII. Social Proximity to Womply and BlueAcorn, Excluding Capital Plus

This table replicates Columns (2) to (5) of Table 6 with loans originated by Capital Plus excluded from the calculation of both the dependent and independent variables. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the number of PPP loans in each zip code. All variables (both independent and dependent) are standardized to have a mean of 0 and a standard deviation of 1. Zip codes are filtered to those with at least 25 loans. Fixed effects and control variables are as indicated at the bottom of each column. Robust standard errors are clustered by county.

	(1)	(2)	(3)	(4)
Dep. Variable:	Womply Per Capita		BlueAcorn Per Capita	
Social Proximity	0.947***	0.883***		-0.184***
to Womply	(14.09)	(5.54)		(-2.88)
Social Proximity		0.0854	1.088***	1.263***
to BlueAcorn		(0.53)	(11.50)	(12.44)
County FE	Yes	Yes	Yes	Yes
Loans Per Capita	Yes	Yes	Yes	Yes
Demographic Controls	Yes	Yes	Yes	Yes
Pct. Friend 50 & 150 Mi	Yes	Yes	Yes	Yes
Observations	19,367	19,367	19,367	19,367
Num. Counties	2,320	2,320	2,320	2,320
R^2	0.878	0.879	0.908	0.909

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table IA.XXIII. Social and Physical Proximity to Lenders

This table examines the effects of physical proximity to FinTech loans and loans originated by different lenders/platforms on borrowers using a FinTech lender or specific lender/platform. Panel A examines the effects of physical proximity. Panel B examines the effects of social and physical proximity when included together. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the number of PPP loans in each zip code. All variables (both independent and dependent) are standardized to have a mean of 0 and a standard deviation of 1. Zip codes are filtered to those with at least 25 loans. Fixed effects and control variables are as indicated at the bottom of each column. Robust standard errors are clustered by county.

Panel A. Effects of Physical Proximity

Dep. Variable:	(1) FinTech Per Capita	(2) Womply Per Capita	(3) BlueAcorn Per Capita	(4) BoA Per Capita	(5) JPMChase Per Capita
Physical Proximity to Lender(s)	0.302*	0.545*	0.881***	0.0860	0.241
	(1.78)	(1.88)	(6.13)	(1.64)	(1.30)
County FE	Yes	Yes	Yes	Yes	Yes
Loans Per Capita	Yes	Yes	Yes	Yes	Yes
Demographic Controls	Yes	Yes	Yes	Yes	Yes
Pct. Friend 50 & 150 Mi	Yes	Yes	Yes	Yes	Yes
Observations	19,367	19,367	19,367	19,367	19,367
Num. Counties	2,320	2,320	2,320	2,320	2,320
R^2	0.868	0.810	0.838	0.814	0.741

Panel B. Effects of Social and Physical Proximity, Included Together

Dep. Variable:	(1) FinTech Per Capita	(2) Womply Per Capita	(3) BlueAcorn Per Capita	(4) BoA Per Capita	(5) JPMChase Per Capita
Social Proximity to Lender(s)	0.849***	1.058***	1.162***	-0.146	-0.164
	(15.51)	(25.53)	(44.49)	(-1.74)	(-1.25)
Physical Proximity to Lender(s)	-0.127***	-0.102**	-0.149	0.137	0.287
	(-4.11)	(-2.02)	(-1.28)	(1.67)	(1.59)
County FE	Yes	Yes	Yes	Yes	Yes
Loans Per Capita	Yes	Yes	Yes	Yes	Yes
Demographic Controls	Yes	Yes	Yes	Yes	Yes
Pct. Friend 50 & 150 Mi	Yes	Yes	Yes	Yes	Yes
Observations	19,367	19,367	19,367	19,367	19,367
Num. Counties	2,320	2,320	2,320	2,320	2,320
R^2	0.934	0.898	0.917	0.816	0.743

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table IA.XXIV. EIDL Agents

This table examines the role of agents in the EIDL program. Panel A examines whether social networks influence agent relationships. Panel B examines whether the flag rate on previous loans facilitated by an agent is predictive of the likelihood that a subsequent loan facilitated by the agent is flagged. *Agent Connectedness* is defined as the number of agents facilitating loans in both zip codes i and j normalized by the product of agents facilitating loans in zip code i and agents facilitating loans in zip code j. *Social Connectedness* is the Facebook Social Connectedness Index between the zip codes as described in Section 4. Loans are considered flagged if the loan is flagged by at least one of the business registry, multiple loans, or EIDL > PPP jobs flags (see Griffin, Kruger, and Mahajan (2023) for details on these flags). Fixed effects and control variables are as indicated at the bottom of each column. Robust standard errors are double clustered by zip code i and zip code j for Panel A and clustered by agent in Panel B.

Panel A. Agent Network

Dep. Variable: Agent Connectedness

	(1)	(2)	(3)	(4)
Zip Code Pairs:	All	$\geq$ 100 Miles Apart	$\geq$ 250 Miles Apart	$\geq$ 500 Miles Apart
Social Connectedness	1.137***	1.549***	1.337***	1.124***
	(22.05)	(9.14)	(6.27)	(4.52)
Zip Code i FE	Yes	Yes	Yes	Yes
Zip Code j FE	Yes	Yes	Yes	Yes
Observations	139,119,538	135,790,680	126,521,888	105,710,156
R^2	0.148	0.149	0.151	0.153
Mean of Dep. Var.	0.00407	0.00405	0.00406	0.00403

Panel B. Effect of Past Agent Flag Rate on Likelihood Loan is Flagged

Dep. Variable: 1(Flagged)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Num. Past Loans:	$\geq$ 1	$\geq$ 5	$\geq$ 10	$\geq$ 25	$\geq$ 1	$\geq$ 5	$\geq$ 10	$\geq$ 25
Agent Flag Rate	0.348***	0.508***	0.542***	0.619***	0.302***	0.433***	0.460***	0.538***
	(39.33)	(24.69)	(16.89)	(16.37)	(38.80)	(25.43)	(16.80)	(12.37)
Zip Code FE	No	No	No	No	Yes	Yes	Yes	Yes
Observations	197,877	102,929	73,997	50,259	193,174	98,834	69,939	46,224
R^2	0.0599	0.0446	0.0288	0.0181	0.156	0.173	0.179	0.204
Mean of Dep. Var.	0.0598	0.0546	0.0502	0.0460	0.0603	0.0552	0.0511	0.0469

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010