NBER WORKING PAPER SERIES

ENTREPRENEURIAL SPAWNING FROM REMOTE WORK

Alan Kwan
Ben Matthies
Richard R. Townsend
Ting Xu

Working Paper 33774
http://www.nber.org/papers/w33774

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2025

We thank Tania Babina, Jose Maria Barrero, Jun Chen, Fabrizio Core, Isaac Hacamo, Yael Hochberg, Vikram Nanda, Toomas Laarits, Fangyuan Ma, Adair Morse, Matthew Pecenco, Kaushik Vasudevan, and conference and seminar participants at the ITAM Finance Conference, HEC Paris Entrepreneurship Workshop, TAMU Young Scholar in Finance Consortium, CUHK-RCFS-RAPS Conference, Winter Finance Summit in Asia, MFA, CMU Tepper, Alberta, CKGSB, Notre Dame, KAIST/Korea University, Univ. of Ottawa, and HKU (Applied Econ) for helpful comments. We thank the anonymous data provider for support. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w33774

Entrepreneurial Spawning from Remote Work
Alan Kwan, Ben Matthies, Richard R. Townsend, and Ting Xu
NBER Working Paper No. 33774
May 2025
JEL No. E32, J22, J24, L26, M13

ABSTRACT

Using a novel firm-level remote work measure created from big data on Internet activity, we show that firms with higher remote work during the pandemic are more likely to see their employees becoming entrepreneurs. This effect holds both unconditionally and relative to other types of job turnovers. We establish causality using instrumental variables and a panel event study. The marginally created businesses are higher quality than the average new firm. The effect is not driven by employee selection, preference change, or forced turnover. Rather, remote work increases spawning by providing the time and downside protection needed for entrepreneurial experimentation.

Alan Kwan
Hong Kong University
apkwan@hku.hk

Ben Matthies
University of Notre Dame
bmatthie@nd.edu

Richard R. Townsend
University of California, San Diego
Rady School of Management
and NBER
rrtownsend@ucsd.edu

Ting Xu
University of Toronto
Rotman School of Management
and NBER
tingxu.xu@rotman.utoronto.ca

"*While people have always worked nights and weekends to start their own businesses, remote work gives them more time and flexibility to do so and a better hedge against failure.*"

– Vox, "The Rise of the Side Startup," 08/11/2022

1 Introduction

The labor market has experienced a massive shift to remote work in the past few years, catalyzed by the pandemic. In 2023, full days worked from home account for 28% of paid workdays in the US, four times higher than the level in 2019. Even in 2025, remote work persisted at 25% of workdays (Barrero et al., 2023; Buckman et al., 2025).[1] At the same time, business formation surged during the pandemic and has stayed high (Decker and Haltiwanger, 2023). This paper examines whether there is a link between these two macro phenomena by testing whether remote work increases workers' transitions to entrepreneurship.

Understanding whether and how remote work spawns entrepreneurship is important because most entrepreneurs transition from wage employment. Hence, frictions within wage employment can impact labor flows to entrepreneurship and, ultimately, growth and innovation (King and Levine, 1993; Decker et al., 2014). There is also growing concern that remote work may inhibit innovation and idea generation (Brucks and Levav, 2022; Lin et al., 2023; Chen et al., 2022). As companies and policy makers continue to evaluate the merits of remote work, its spillover effect on entrepreneurship could be an important consideration in their cost-benefit calculations.

Answering the above question, however, is empirically challenging. First, we need to be able to accurately measure remote work at the firm level. Yet most available measures are survey-based or cover only a limited set of firms or jobs. Second, we need to observe worker-level transitions into entrepreneurship and worker-firm matches. Finally, variation in remote work adoption across firms is not random. Firms that adopt more remote-work-friendly policies may already employ more entrepreneurial individuals, making it difficult to establish causality.

To overcome the measurement challenge, we exploit big data on Internet activities to create a firm-level measure of remote work. The data allow us to classify IP addresses and track anonymized individuals across their workplace, home, and mobile devices. It also links individuals to their employers. We aggregate these data across all employees of a firm to obtain a firm-month-level measure of remote work—the percentage of employee Internet activities during work hours that belong to a remote IP address. By measuring firm-wide remote work rather than individual take-up

[1] Similar patterns hold in other countries (Buckman et al., 2025).

of remote work, we mitigate individual-level endogeneity, exploiting the fact that individuals have limited influence over firm-wide policies. As such, we estimate an "intent-to-treat" effect. To validate our remote work measure, we compare it with SafeGraph mobility data, telework potentials of jobs, and mentions of remote work in job postings. We find that aggregate remote IP traffic during work hours increased significantly in March 2020, while remaining stable during non-work hours.

We then link our firm-level remote work measure to employer-employee matched data from LinkedIn, available through Revelio Labs. The LinkedIn data contain the job history of each worker. This allows us to observe transitions into entrepreneurship from wage employment, i.e., spawning. We also observe the characteristics of workers and their employers. To mitigate potential truncation issues from stale resumes on LinkedIn, we track spawning activities up until December 2022, even though our LinkedIn data end in October 2023. Although not all workers are on LinkedIn, our data capture the set of knowledge workers at risk of becoming an entrepreneur—the same set that is also well covered by our Internet activity data.

Our baseline cross-sectional analysis focuses on U.S. firms with at least 10 employees in February 2020 ("Feb2020 firms"), the month before COVID, and all their employees as of then ("Feb2020 employees").[2] We examine the effect of a firm's average remote work in 2020/21 on the spawning activities of its Feb2020 employees from March 2020 to December 2022. We conduct this analysis at both the individual and the firm level. Our baseline OLS estimates show that a one-standard-deviation increase in remote work is associated with a 5% higher likelihood of entrepreneurial spawning at the individual level relative to the mean, and a 4% higher spawning share among employees at the firm level relative to the mean.

We take several approaches to address the potential endogeneity of remote work policies. First, we control for a host of ex-ante firm and employee characteristics in our baseline specification, including firms' pre-pandemic remote work level and spawning share, as well as workers' past founder experience. These controls help absorb ex-ante entrepreneurial tendencies at both the firm and worker level, such as firm culture or employee types.

Second, we use instrumental variables to isolate exogenous variation in remote work. Our primary instrument is the average commute distance of a firm's employees before the pandemic, when remote work was not a consideration for most firms. The idea is that firms whose workers lived farther away from the office before the pandemic will be more likely to adopt remote work during and post pandemic. We posit that within a given location (captured by county fixed effects) and conditional on pre-pandemic remote work, residual variation in commute distance is largely idiosyncratic and predetermined before the pandemic, thus providing an exogenous source of variation in

[2] We exclude firms with fewer than 10 employees as our firm-level remote work measure is less accurate for these firms.

remote work tendencies. We also use local business closure orders issued during the pandemic as an alternative instrument. We verify that these instruments have no direct correlation with firms' pre-pandemic spawning in the cross-section conditional on our controls, and that firms sorted by these instruments trended similarly before the pandemic. Our preferred 2SLS estimate shows that a one-standard-deviation increase in a firm's remote work increases the spawning share among its employees by 42% relative to the mean.

Third, we use a dynamic difference-in-differences (DID) approach to compare changes in firms' annual spawning rates around the pandemic across firms with different tendencies to adopt remote work after 2020, as proxied by our instruments. To remove employee selection effects, we track a fixed set of employees employed in February 2020 over time, regardless of whether they were with their Feb2020 employer in a year. This sample shuts down employee recomposition and hence the possibility that our results are driven by selection—e.g., remote work increase attracting more entrepreneurial employees or driving away less-entrepreneurial ones. This DID analysis yields results consistent with our cross-sectional results. We also use this specification to examine one's "entrepreneur status," i.e., whether spawned individuals remain in entrepreneurship after spawning. This allows us to test if spawned individuals quickly returned to wage employment. We find a persistent and increasing effect on this outcome from 2020 to 2022, suggesting that the spawned entrepreneurship is not transient.

Our baseline results are robust to alternative samples, additional controls, and alternative definitions of spawning and remote work. Our main analysis excludes firms with more than 5000 employees to mitigate individual-level results being skewed by the largest firms. We show that our results are robust to including them. We also show that our results are robust to including additional controls such as workers' age, education, and job role. Our results are similar when restricting to spawning events that happened after an individual leaves a wage job, suggesting remote work is not only spawning part-time, side entrepreneurship. We find similar results measuring remote work (RW) from residential IP traffic only (i.e., excluding mobile and VPN). To further tighten identification and remove firm unobservables, we compare workers working for the same firm but in different establishment locations, whose remote work policies were influenced by local social distancing measures during COVID. We find similar results in this within-firm comparison.

To rule out that our results reflect COVID-specific demand shocks, we show that our results are robust to dropping the top booming industries during COVID, including industry-year fixed effects in our dynamic DID, and controlling for worker job role fixed effects. Our results are not driven by the remotability of the spawned firm, as they are robust to comparing workers in the same location working for different firms; we also show that the marginally spawned firms are just as likely to be remote as in-person. Finally, using Census data, we show that our micro-level evidence also holds at the aggregate level: industries or locations with more remotable jobs saw greater new employer firm entry post the pandemic. This analysis mitigates any concerns with our LinkedIn- and Internet-based measures.

To what extent are our findings unique to entrepreneurship? It is possible that forces that drive remote workers into entrepreneurship also drive them to other wage employers or non-employment. To investigate this, we conduct a conditional analysis, restricting our sample to employees who experienced job turnovers, i.e., those who left their Feb2020 employer between March 2020 and December 2022. Examining entrepreneurial spawning among this sample thus tests whether remote work *disproportionately* directs workers to entrepreneurship relative to other labor destinations. We find similar results with this conditional analysis. A one-standard-deviation increase in remote work increases spawning into entrepreneurship relative to other job turnovers by 30% under 2SLS. As such, our results do not just capture a general turnover effect of remote work; rather, remote work uniquely shifts workers toward entrepreneurship. The conditional result also rules out any remaining concerns about truncation bias from stale resumes, as we condition our analysis on observing a resume update. It also rules out unobserved demand-side shocks, as such shocks should explain labor reallocations in general, not just entry into entrepreneurship.

We explore three non-mutually exclusive mechanisms for our results: 1) preference change, 2) experimentation, and 3) forced entrepreneurship. Under preference change, remote work induces a preference towards flexibility or "a quiet life" with less employer monitoring. If such a preference drives our results, we should expect the marginal entrepreneurs to concentrate in low-growth, flexibility-based entrepreneurship, such as self-employment. However, we find that the marginally spawned firm is more likely to be an employer and receive subsequent VC funding than the average new firm. This suggests that the marginal entrepreneur tends to be high-quality. Alternatively, prolonged exposure to remote work can induce a yearning for in-person interactions, motivating workers to start their own businesses to fulfill their social needs. This should predict that the marginally spawned firm is more likely to operate in person. However, we find no such evidence: marginally spawned firms are just as likely to have high levels of remote work as low levels.

The experimentation channel posits that remote work spawns entrepreneurs by providing the time and downside protection needed for entrepreneurial experimentation (Kerr et al., 2014). Remote work frees up time by removing commute, increasing productivity, and offering more flexible hours. Such slack time can be used by a worker to develop and experiment with business ideas (Agrawal et al., 2018). Remote work also offers less employer monitoring, which helps to keep a worker's side exploration in "stealth," reducing downside career risks. All these allow a worker to better use her wage employment as a fallback option when exploring entrepreneurship (Gottlieb et al., 2022). If this channel is at work, we should see stronger marginal entry into industries where experimentation is more valuable, such as those with a higher risk of failure. We indeed find such heterogeneity when splitting spawning by the average young firm failure rate in the destination industry. Further, if remote work enables experimentation by relaxing time constraints, we should expect our results to concentrate where such constraints are more binding: e.g., high-growth entrepreneurship that is more time-consuming; our prior result on entrepreneurship quality supports this prediction. Exploiting variation in slack time from child care, we further show that the spawning response is stronger when local K12 schools adopted more in-person learning during the pandemic, which gave remote work parents more time to experiment with entrepreneurship.

Finally, our results could reflect forced entrepreneurship, where remote work leads to layoffs or involuntary departures, and these workers subsequently start a business out of necessity. We rule this out by showing that our findings are similar when restricting to firms that experienced continued employment growth from 2020 to 2022, i.e., those unlikely to have had mass layoffs. Overall, our evidence is most consistent with remote work spawning entrepreneurship by providing workers the time and protection needed for entrepreneurial experimentation.

We end our paper with a back-of-the-envelope calibration at the macro level. Based on our firm-level estimate, we calibrate that at least 11.6% of the post-pandemic increase in new firm entry can be explained by spawning from remote work. Of course, there could be other channels through which remote work impacts entrepreneurship at the aggregate level, such as investment opportunities or local agglomeration. Nevertheless, our paper uncovers a novel link between two important macro phenomena post-pandemic: the rise of remote work and increases in business entry.

Our findings highlight mechanisms that would likely persist beyond the pandemic context, suggesting broader external validity. The channels through which remote work enables entrepreneurship—increased protection for experimentation from reduced employer monitoring and greater time availability—should hold true for any level of remote work, though with varying intensity. Hybrid arrangements would likely have a weaker effect than fully remote models due to more regular in-person monitoring and less commute time saved. Importantly, remote work represents a new equilibrium rather than a temporary pandemic response, with Barrero et al. (2023) showing that remote work has persisted well beyond the pandemic and is predicted to remain a permanent feature of the labor market. However, we acknowledge that the COVID period was exceptional in ways that may have amplified our observed effects, particularly the availability of government subsidies and the abundance of VC funding. In environments with more constrained capital markets or without government support programs, the magnitude of remote work's impact on entrepreneurial spawning might be less pronounced, even if the underlying mechanisms remain valid. Still, our findings help explain the notable rise in entrepreneurship during this period.

Our paper adds to the fast-growing literature on remote work (see Barrero et al. (2023) for a review). The literature has shown that remote work persisted after the pandemic and is predicted to stay in the long run, due to both learning and better remote work technologies (Barrero et al., 2021b; Aksoy et al., 2022). There is large variation in the adoption of remote work across occupations, geographies, firms, and industries (Hansen et al., 2023; Aksoy et al., 2022). The productivity impact of remote work is largely positive for hybrid arrangements, but is mixed for fully remote (Bloom et al., 2015, 2024; Kwan et al., 2023; Emanuel and Harrington, 2023; Gibbs et al., 2023; Duchin and Sosyura, 2021; Flynn et al., 2024; Atkin et al., 2023), with more negative effects on innovation (Brucks and Levav, 2022; Lin et al., 2023; Chen et al., 2022). We focus on the spillover effect of remote work on entrepreneurship. Our findings suggest that the impact of remote work on aggregate productivity may be higher than firm-level estimates.

In closely related contemporaneous work, Barrios et al. (2024) also examine the relationship between remote work and entrepreneurship. They find that while overall entrepreneurial entry increased with the onset of the pandemic, areas with more remotable jobs prior to the pandemic saw a relative decline in new business registrations during the pandemic, suggesting that remote work substitutes for entrepreneurship. Our paper differs in that we conduct our analysis at the individual and firm levels, rather than at the zipcode level. In addition, our measure of entrepreneurship is based on founder status on LinkedIn rather than business registrations. This LinkedIn-based measure tends to capture more growth-oriented ventures compared to business registrations, which predominantly consist of small businesses. While Barrios et al. (2024) emphasize how increased workplace flexibility in traditional employment may suppress small business entrepreneurship motivated by nonpecuniary benefits such as autonomy and flexibility, our paper complements their findings by showing that for high-growth entrepreneurs, remote work can actively encourage entrepreneurial entry by providing workers with additional time and reduced risk to experiment with entrepreneurship.

We also contribute to the literature on labor and entrepreneurship. Babina (2020) and Babina and Howell (2024) document entrepreneurial spawning from the financial distress and R&D of incumbent firms. Gompers et al. (2005) show that the most prolific spawners are originally VC-backed firms. Hacamo and Kleiner (2022) and Bernstein et al. (2024) examine how economic cycles affect the decision to become an entrepreneur or work for a startup firm. Hombert et al. (2020) study how downside insurance for unemployed workers affects the quality of new firms started by these individuals. Our paper examines how a new paradigm of work impacts entrepreneurship. We show that remote work provides the safe space needed by workers to experiment with entrepreneurial ideas before formally launching a business. As such, our mechanism is related to Gottlieb et al. (2022) and Barrios et al. (2022), who respectively show that job-protected leave and gig work opportunities encourage entrepreneurial entry by providing a safety net for experimentation.

Finally, our study has implications for the spatial distribution of entrepreneurship and innovation, which is known to be highly concentrated in a few hub cities (Feldman, 2013; Florida and Mellander, 2016). Because remote work removes geographic barriers, it results in a more dispersed geographic distribution of workers (Brueckner et al., 2023; Ramani and Bloom, 2021) and, by our paper, distribution of startups spawned by them. As such, remote work could be an important policy tool to reduce spatial inequalities in entrepreneurship (Glaeser and Hausman, 2020).

2 Data, Measures, and Samples

2.1 Data

Internet Activity Data. We create our firm-level remote work measure from novel Internet activity data. The data consist of individual-level Internet activities, including the user, the IP address, and timestamps of access. The data also include information about each user such as device type, approximate latitude and longitude when accessing the Internet, and the company they worked for. The data capture a substantial fraction of Internet activities, comprising approximately one-fifth of the 4 billion IPv4 addresses in the world. Since websites, as well as some servers and Internet-connected devices, are assigned IP addresses, the IP addresses we observe likely comprise an even larger fraction of IP addresses primarily used for human content consumption.

We obtain the data through a partnership with a data analytics company from the marketing technology space, the "Data Partner." The Data Partner maintains a large network of partnerships with online publishers, focusing primarily (but not exclusively) on business content and news. Contributors include thousands of major Internet publishers. Most participate anonymously but span a wide range of business functions such as technology, finance, marketing, legal, human resources, manufacturing, science, and general business. Participating publishers contribute to the Data Partner's pooled dataset via a technology mechanism that shares information about web content consumption, including the external IP address of the network originating the HTTP request. Overall, the platform aggregates around 1 billion content consumption events per day. From this dataset, the Data Partner performs two steps: (1) associating visitors with the companies they work for, when possible, and (2) quantifying the "topics" of the content visitors read about.[3] Our access to the data is designed to take special care with respect to confidentiality restrictions—while we

[3]From these two steps, the Data Partner produces analytics that are primarily sold to companies to facilitate sales and marketing: by identifying companies with heightened research interest in a specific business topic, these companies can target potential customers. Participating publishers receive some of the data analytics in return for providing the data.

observe browsing activity at the event level, we do not know the identity of any individuals in the data.

To associate users with the firms they work for, the Data Partner creates a profile through the use of first- and third-party cookies. This enables the publisher, and in turn the Data Partner, to observe when a visitor returns to a website. Over time, the Data Partner infers the association between the profile and their place of employment (Company) through a wide ensemble of industry-accepted methods. For example, user profiles are associated with a Company when visitors use a work email to log into a participating publisher's website. Another example is through IP addresses. If a profile consistently logs onto a publisher's website from a work-associated IP address, this gives a strong indication that the profile belongs to a particular company. The Data Partner also receives data from third-party sources who perform identity resolution of visitors. Through its proprietary processes, the Data Partner determines whether a reliable association between a profile and a company can be inferred.

Crucially, once a visitor has been associated with a company reliably, the visitor is associated with that company even though the visitor may traverse different IP addresses. This is the primary mechanism through which we are able to monitor transitions between different types of IP addresses, and thus whether the employee is remote or not.

A notable limitation of the data is that one can only observe Internet activities in the cooperative. In addition, mappings between users and employers and between IP addresses and locations are imperfect. However, we hope given the large magnitude of available data, idiosyncratic noise in linking individuals to employers or classifying IP addresses can be mitigated. We also impose sample filters to reduce noises in using the data to measure remote work (see Section 2.2). It is worth noting that any measurement error should induce an attenuation bias in OLS, making it harder for us to find a significant effect. However, such bias should be corrected in 2SLS.

Employer-Employee Matched Data. We obtain employer-employee matched data from Revelio Labs, which is based on LinkedIn data. Our data consist of the universe of LinkedIn users, their resumes, and their employer profile pages up to October 2023. The resume data include each individual's job history, education, skills, demographics, and other information. Revelio/LinkedIn is used by many economics and finance studies (e.g., Agrawal et al. (2021), Chen et al. (2023), and Eisfeldt et al. (2023)), including studies on entrepreneurship (e.g., Hacamo and Kleiner (2022), Jeffers (2024), Bernstein et al. (2024)). This data also give us firm-level employment size, industry, business description, founding year, and website (if any).

One limitation of LinkedIn data is that not all workers are on LinkedIn. However, LinkedIn likely captures the set of workers we are interested in, i.e., those *at risk of* spawning. These tend to be knowledge workers. Additionally, workers on LinkedIn overlap well with those tracked by our Internet activity data, as both sets tend to have an online presence. Another limitation is the potential truncation issues associated with stale LinkedIn profiles. We discuss how we address truncation concerns in Sections 2.2 and 5.

Other Firmographic Data. We supplement Revelio/LinkedIn data with other firmographic databases such as Aberdeen CiTDB and People Data Labs (PDL), which source firm profiles from various sources. These data give us additional information on firms' NAICS codes, locations, founding years, domains, etc.

US Census Aggregate-Level Data. Finally, we use industry- or county-level new firm entry data from Business Dynamic Statistics (BDS) and new firm job creation data from Quarterly Workforce Indicators (QWI) to verify our micro-level results at the aggregate level.

2.2 Key Measures

Remote Work Measure. Our firm-level remote work measure, RW, comes from Kwan et al. (2023). The measure is premised on classifying IP addresses using pre-pandemic data. The classification covers over 760 million IPs, about 20% of possible IPv4s and a likely greater fraction of IPv4s used by *humans* (a large number of IP addresses belong to servers). The IPs are classified into one of four categories: business, residential, mobile, or VPN. The classification is conducted using a two-step approach: first using rules-of-thumb to classify IP addresses, and then using a machine learning model to categorize the remaining unclassified IPs. Importantly, our classification is based on pre-pandemic information. Appendix A.1 provides details on our IP classification.

To compute the extent to which a firm is working remotely, we calculate the fraction of the firm's IP traffic during work hours (Mon-Fri 9am-5pm) that is originating from a remote IP. We define a remote IP as a residential, mobile, or VPN IP — that is, any IP that is not an office or business address. We compute this fraction at the firm-month and firm-year level.[4] Our RW measure is available from 2019 to 2021. The RW measure can also be flexibly constructed at the local or industry level.

[4]We restrict to firms for which we can reliably measure RW. In particular, we restrict to firm-months satisfying the following criteria: 1) in a given month, have 100 work-time observations, where an observation is a session-timestamp from an employee of the firm; (2) average at least 1,000 observations per month for at least half of the months from January 2019 to February 2020, and half of the months from March 2020 to December 2021; (3) are included in one of our firmographic databases.

Because users are anonymous in our Internet data and cannot be linked to LinkedIn employees, we are not able to measure remote work at the employee level. However, the benefit of a firm-level measure is that it is more exogenous than an individual-level measure, as an individual employee has limited influence over firm-wide policies. As such, we can think of individual-level RW as "take-up," and firm-level RW as "intent-to-treat." Given a firm's RW policy, an individual employee's decision to take up is obviously more endogenous, as it is driven by the person's expected costs and benefits of take-up, which could correlate with her unobserved entrepreneurial tendencies.

Figure 1 shows the time series of monthly *RW* averaged across firms in our sample, with January 2019 normalized to zero. Remote work increased sharply at the start of the pandemic and stayed elevated, with a slight decline in 2021. Relative to the survey-based measure from Barrero et al. (2023), we see less of a reversal in 2021. This likely reflects the fact that, among the population of knowledge workers we capture, remote work is more persistent, whereas Barrero et al. (2023) survey the general population. Table A.1 shows the top industries with the most or least increase in remote work relative to pre-pandemic. As expected, IT and professional services had the highest increases in *RW*, while retail trade, construction, and agriculture had the lowest.

The level of *RW* may capture some differences in the ways companies manage their networks. For example, the pre-pandemic value of *RW* is not close to zero, suggesting there is a baseline level of non-office IP activities among in-office workers.[5] This arises for two reasons. First, Internet traffic from workers who are home sick, on vacation, traveling for work, or working non-standard hours may be attributed to remote activity. Second, idiosyncratic technical reasons may lead to over-classification of remote work, such as using cell phones in the office while on a mobile network (i.e., not through the firm's WiFi), or taking morning emails at home or at a coffee shop. Importantly, many of these features vary at the industry level, which our industry fixed effects absorb throughout. Additionally, all our analyses either control for pre-pandemic *RW* or focus on within-firm changes in *RW*. This also ensures that we do not capture any pre-pandemic differences in remote work across firms, which may correlate with corporate culture or telework infrastructure, etc. In Section 4.3, we demonstrate robustness to measuring *RW* using residential IP traffic only (i.e., excluding mobile and VPN traffic).

We conduct extensive validations of our remote work measure. First, remote IP activity during work hours significantly increased after the onset of the pandemic, but remains largely flat when measured during non-work hours (Figure A.1). Second, at the county-week level, remote traffic *during the day* increases when SafeGraph reports people going into the office less, with an elasticity of -76% (Table A.2). This elasticity drops significantly during the night. Third, *RW* correlates strongly with county-level business closure orders imposed during COVID (Figure A.2). Fourth, at

[5] Surveyed remote work share was 7% in 2019 (Bloom et al., 2024).

both the MSA and industry level, our measure correlates strongly with the ex-ante remotability of jobs (i.e., telework potential) from Dingel and Neiman (2020) (Figure A.3). Last, at the *firm level*, our measure correlates strongly with realized remote work mentioned in job postings (Figure 2).[6]

Spawning. We use LinkedIn employment history to measure spawning from wage employment into entrepreneurship. A spawning event is defined as an individual reporting a new position with the following conditions met *simultaneously*:

1. The new position is with a company different from the prior employer
2. The job title contains "founder" (including "co-founder"), "founding," "owner," or "entrepreneur." In the case no employees of the firm have any such titles, we use the titles "CEO," "partner," or "president"
3. The individual is within the first five employees of the new firm ranked by job start date
4. The job start date is within one year of the firm's official founding date (reported by LinkedIn)

Although our LinkedIn data is as of October 2023, we track spawning events until December 2022 to mitigate potential truncation bias from stale resumes. We use the new job start date as the spawning event date. In some cases, spawning happens before a person formally leaves her salaried job. Such overlap can happen either because side entrepreneurship is permitted by the employer, or because the founder retroactively reports the new firm start date after she quits and the startup gets out of the "stealth" mode.

For our cross-sectional sample, we track all spawnings from the Feb2020 firms from March 2020 to December 2022. For our firm-year-level panel, we track spawning events for each firm-year. Because spawning is low frequency, we multiply both the individual-level spawning indicator and the firm-level spawning share (i.e., fraction of employees that spawned) by 100, for ease of interpreting coefficients.

2.3 Samples

Our firm-level cross-sectional sample starts with all firms on LinkedIn with at least one employee as of February 2020 with non-missing RW measure. We refer to these firms as "Feb2020 firms." To make sure we can reliably measure firm-level RW, we restrict to firms with at least 10 employees

[6]We follow Hansen et al. (2023) and construct firm-level measure of remote job postings using machine learning. Appendix A.3 details this process and why our IP-based measure is superior.

as of February 2020.[7] This also ensures that we capture entrepreneurial spawning from relatively established, "incumbent" firms. Our individual-level cross-sectional sample consists of all US-based employees of these Feb2020 firms as of February 2020. We refer to this sample as "Feb2020 employees." To mitigate the concern that our individual-level sample is skewed by mega firms, we exclude firms with more than 5000 employees. Our results are robust to including these firms. Our baseline cross-sectional sample has about 13.5 million workers from 136,121 firms.

We also construct a firm-year-level panel covering the period of 2016 to 2022. The panel tracks the Feb2020 employees over time regardless of whether they were with the Feb2020 firm in a given year. Fixing the composition of employees over time helps us remove the effect from employee selection. Section 3 provides more details on why and how we construct this sample.

2.4 Summary Statistics

Table 1 provides summary statistics for our cross-sectional samples. The mean spawning rate from March 2020 to Dec 2022 was 0.34% across all employees. At the firm level, the average spawning share over the same period was 0.43%. The difference reflects the fact that larger firms tend to have lower spawning rates (Gompers et al., 2005). The average firm-level remote work during 2020/21—RW_{post}—is 0.69. There is also much cross-sectional variation in RW_{post} across firms, as shown in the histogram in Figure 3. The median firm in our sample has 27 employees and is 35 years old. The median employee in our sample has a job tenure of 3 years as of 2020, holds a junior rank (seniority=2), and has a salary of approximately $72,000. About 0.2% of the employees have prior founding experience. Table 2 presents the summary statistics for our firm-year-level panel. The mean spawning rate is lower than in the cross-sectional sample both because we measure the annual spawning rate instead of the cumulative spawning rate, and because pre-pandemic spawning rates are lower than post-pandemic.

3 Empirical Strategies

3.1 Cross-Sectional Analysis

Individual Level. Our individual-level analysis focuses on a cross-section of workers employed as of February 2020. We then track these individuals' spawning activities from March 2020 to December 2022, and relate this outcome to their Feb2020 employer's RW in 2020-2021. Specifically,

[7] Our results are similar if restricting to firms with at least 5 or 20 employees.

we estimate the following specification:

$$Spawn_{post,i} = \alpha_n + \beta_c + \gamma RW_{post,f} + \lambda \boldsymbol{X_i} + \rho \boldsymbol{X_f} + \epsilon_i \tag{1}$$

In this equation, the dependent variable $Spawn_{post,i}$ is 100 times a dummy that equals one if the worker i employed with firm f in February 2020 subsequently started a new business between March 2020 and December 2022. The key independent variable $RW_{post,f}$ is the Feb2020 employer's average RW in 2020 and 2021.

We include fixed effects for the Feb2020 firm's NAICS 4-digit industry (α_n) and headquarter county (β_c). We also include a host of individual- and firm-level controls. $\boldsymbol{X_i}$ is a vector of individual-level controls that include job tenure, seniority, log salary, and an indicator for past founder experience, all measured as of February 2020.[8] $\boldsymbol{X_f}$ is a vector of firm-level controls that include the firm's RW in 2019 (RW_{pre}), log employment size, firm age, and its lagged entrepreneurial spawning rate, all measured in 2019. Importantly, controlling for individuals' past founder experience and firms' past spawning rates helps absorb time-invariant unobservables that may affect spawning at both the individual and the firm levels, such as firm culture or employee types. We cluster standard errors by firm industry (NAICS 4-digit).

As a robustness, we further include several additional individual-level controls: age, education, and fixed effects for job roles in February 2020. We infer a person's age based on his/her undergraduate degree year, and in case it is missing, high school completion year. For education, we control for whether the individual has a graduate degree and whether her undergraduate degree is from a top-100 school based on the Times higher education ranking. Since 32% of the individuals do not report any education information, we dummy out these individuals.

To mitigate the concern that our estimated individual-level effect is skewed by the largest firms, we restrict to firms with no more than 5000 employees as of February 2020. Our subsequent firm-level analysis also addresses this concern by weighting each firm equally. We show robustness to relaxing this restriction.

Firm Level. Our firm-level specification is analogous to the individual-level, except that we collapse all individual-level variables to firm-level averages. As such, our dependent variable is the share of Feb2020 employees who started a business between March 2020 and December 2022, and our firm-level controls $\boldsymbol{X_f}$ now include employees' average job tenure, seniority, log salary, and past founder experience as of February 2020, in addition to the ones in Equation 1. This yields the

[8] Salary is inferred by Revelio Labs based on job location, role, seniority, etc.

following firm-level specification:

$$SpawnShare_{post,f} = \alpha_n + \beta_c + \gamma RW_{post,f} + \rho \boldsymbol{X_f} + \epsilon_f \tag{2}$$

2SLS. To isolate exogenous variation in firms' remote work, we also estimate a 2SLS version of the individual-level and firm-level regressions, instrumenting $RW_{post,f}$ with two instruments.

Our primary instrument, $Commute_f$, is the firm-level average commute distance by employees in 2019, calculated using our Internet activity data. The intuition is that firms whose employees lived farther from the office face higher costs of commuting pre-pandemic, when remote work was not a major consideration in firms' decisions. After the onset of the pandemic when remote work first became a consideration for firms, we posit that, all else equal, firms with greater commute distances were more likely to experiment with remote work. For example, consider two identical firms in Manhattan. If one firm's employees lived mainly in Connecticut with a two-hour commute and the other's lived mainly in Manhattan with a 20-minute commute, during the pandemic when both firms were considering remote work policies, the first firm is more likely to implement such policies, which should also persist longer into/after the pandemic. We posit that, within the same location and conditional on our controls, including RW in 2019, residual variation in commute distance across firms is largely idiosyncratic and predetermined before the pandemic. This residual variation could reflect factors such as the proximity of the office to public transit stations or highways, as well as the availability and cost of parking. Thus, the instrument provides an exogenous source of variation in the propensity to work from home.

We construct a measure of commute distance at the firm level for all US firms in our sample using the Internet activity data. We leverage two key features of the data: first, our data are sufficiently granular that we observe the browser session of each individual associated with a firm [9]; second, metadata associated with each IP address allows us to observe the approximate location of each worker whenever they access the Internet. We compare each employee's location during nonworking hours with their location during working hours (Monday through Friday between 9 a.m. and 5 p.m.). This allows us to calculate the approximate commute distance from each employee's home to the office for each firm in the United States.[10] We calculate the haversine distance (in kilometers) between the approximate home location and the office location for each user session in each firm during 2019. For each firm, we compute the 25th and 75th percentiles of the commute distance across all sessions during 2019. We then calculate the average commute distance within

[9] A browser session is an approximation of an employee, but an individual may have multiple browser sessions if they go incognito or use different browsers.

[10] We exclude employees whose work-time and non-work-time locations are exactly the same (i.e., already working from home). We do not observe any personal identifying information about any employee. We also do not observe precise locations of employee residences – we observe only the approximate neighborhood or district of each employee.

the middle 50 percentile of the distribution to obtain $Commute_f$, the average commute distance of employees at firm f.[11] Appendix A.2 provides more details on the construction and validation of this instrument, including validation using SafeGraph data.

There are two potential concerns with using commute distance as an instrument. First, commute distance varies by firm geography. To address this issue, we normalize our commute distance measure within a zipcode. This allows us to compare the commute distance of firms within the same location and ensures that the instrument does not capture urban-suburban differences. We further include county fixed effects in our analysis. Second, commute distance could be correlated with worker or firm characteristics. In our panel analyses, we fix employee composition to directly address this concern. Our cross-sectional analysis controls for pre-pandemic firm industry, size, age, and remote work level, which capture firms' differential ability to implement remote work prior to the pandemic. Additionally, we control for firms' pre-pandemic spawning rates and workers' prior founder experience, which account for employees' underlying entrepreneurial tendencies. These controls greatly reduce the residual correlation between commute distance and unobserved firm types that does not operate through remote work.

Our alternative instrument is based on county-level business closure orders issued during COVID. We obtain local business open and closure orders used in Spiegel and Tookes (2021) from the Tobin Center for Economic Policy and the Yale School of Management. We then compute the average fraction of time over 2020 to 2021 when businesses were required to be closed in a county, taking into account both full and partial closures. Specifically, Spiegel and Tookes (2021) categorize four levels of business open measures: medium risk open, high risk open, higher risk open, and highest risk open. We assign a weight of 50%, 33.3%, 16.7%, and 0% to each of these levels to compute the average closure time. This instrument plausibly satisfies the exclusion condition because these measures were implemented by local politicians, partly in response to the severity of the local pandemic situation. In other words, the instrument isolates "forced" remote work changes. However, the downside is that this instrument is at the county level and requires us to drop county fixed effects.

For our 2SLS to estimate a valid local average treatment effect (LATE), our instruments need to satisfy three assumptions: 1) relevance, 2) monotonicity, and 3) exclusion. We present evidence supporting each assumption in Section 4.1.

[11] This removes the two tails of the distribution, which tend to be ephemeral or outlying sessions that are not accurately measured.

3.2 Firm Panel Analysis

To gauge the dynamics of our cross-sectional results, we also conduct a firm-level panel analysis. We construct this sample from a fixed set of employees employed in February 2020. Relative to examining all employees in each firm-year, our fixed-employee sample removes employee recomposition and hence the possibility that our results reflect employee selection. For example, firms that increased remote work may subsequently attract new employees who are more entrepreneurial, and these employees subsequently became entrepreneurs.[12]

Specifically, we start with a balanced individual-level panel for the Feb2020 employees. We track all their spawning activities from 2016 to 2022, regardless of which firm they spawned from. In other words, we allow their employers to change and be different from their Feb2020 employer. We then link these spawning events to the remote work policies of the Feb2020 firm. By fixing the composition of employees, we remove individuals' selection into Feb2020 firms based on unobserved characteristics. Effectively, this approach differences out individuals' latent spawning tendencies using their spawnings from other employers before or after the Feb2020 employer. We then collapse this individual-level panel to the Feb2020-firm-year level and estimate the following DID specification:

$$SpawnShare_{f,t} = \alpha_f + \beta_t + \theta \cdot Treat_f \cdot Post2020_t + \gamma \cdot \boldsymbol{X_f} \cdot Post2020_t + \epsilon_{f,t} \tag{3}$$

, where f indicates the Feb2020 firm, $Post2020_t$ indicates years$\geq$2020, and $\boldsymbol{X_f}$ are ex-ante firm-level controls used in Equation 1. The dependent variable is the share of the Feb2020 employees that started a new business in a year. $Treat_f$ is a continuous treatment variable that is either ΔRW_f—within-firm change in RW from 2019 to 2020/21 (i.e., $RW_{post} - RW_{pre}$), or one of our two instruments $Commute_f$ and $BizClose_c$. We standardize $Treat_f$ to interpret a one-standard-deviation change effect. This specification thus estimates a generalized DID, in either OLS or reduced form. We include firm fixed effects (α_f) for the Feb2020 firms and calendar year fixed effects (β_t). $\boldsymbol{X_f} Post2020_t$ controls for differential shocks that vary by firm characteristics. As robustness, we additionally control for industry-year fixed effects to absorb sector-specific shocks during COVID. Standard errors are clustered by Feb2020 firms' NAICS 4-digit industry.

We also estimate a dynamic version of the baseline DID, omitting 2019 as the base year:

$$SpawnShare_{f,t} = \alpha_f + \beta_t + \sum_{t \neq 2019}^{2016 \rightarrow 2022} \theta_t \cdot Treat_f \cdot \mathbb{1}(Year = t) + \gamma \cdot \boldsymbol{X_f} \cdot Post2020_t + \epsilon_{f,t} \tag{4}$$

[12] A priori, the selection effect could also go the other way: firms with generous remote work policies retain employees who prefer flexibility or a quiet life, while firms quickly reverting back to in-person lose such employees, who then start a new business for flexibility reasons.

This event study also tests if the parallel trends assumption is likely to hold in our sample, i.e., firms with different levels of ΔRW, $Commute$, or $BizClose$ trended similarly before 2020.

One concern with tracking a fixed set of individuals over time is that spawnings may be mean-reverting. If an individual just left wage employment to start a new firm, it will be hard to observe another spawning event immediately after, given that she needs to return to wage employment before she can spawn again. In other words, within a short period of time, many individuals can only spawn once. Another concern is that the spawning event itself does not tell us how persistent the effect is, i.e., how long the individual stays in entrepreneurship after spawning. It is possible that spawned individuals quickly reverted back to wage employment, either because the started business quickly failed, or because it was a temporary arrangement while individuals were between jobs. In either case, this would suggest remote work only spawns low-quality entrepreneurship.

To mitigate these concerns, we examine one's founder status as an alternative outcome. We define an individual as a founder as long as she is in entrepreneurship in a given year, regardless of whether this is the first or second business she founded after transitioning to entrepreneurship. We then similarly collapse the individual-level panel to the Feb2020-firm-year level to define $FounderShare$. This alternative outcome is not subject to mean reversion. Additionally, it captures how long one stays in entrepreneurship after being exposed to remote work in their initial wage job. We estimate the same specifications as Equations 3 and 4, but with $FounderShare_{f,t}$ as the dependent variable:

$$FounderShare_{f,t} = \alpha_f + \beta_t + \theta \cdot Treat_f \cdot Post2020_t + \gamma \cdot \boldsymbol{X_f} \cdot Post2020_t + \epsilon_{f,t} \tag{5}$$

$$FounderShare_{f,t} = \alpha_f + \beta_t + \sum_{t \neq 2019}^{2016 \rightarrow 2022} \theta_t \cdot Treat_f \cdot \mathbb{1}(Year = t) + \gamma \cdot \boldsymbol{X_f} \cdot Post2020_t + \epsilon_{f,t} \tag{6}$$

4 Main Results

4.1 Cross-Sectional Results

Table 3 presents the cross-sectional OLS result. Column 1 reports the individual-level result based on Equation 1, where the dependent variable is 100 times the spawning indicator. We find that workers who experienced higher levels of remote work during COVID were significantly more likely to leave their employer and start a new business between March 2020 and December 2022. In particular, a one-standard-deviation increase in RW_{post} increases worker-level spawning likelihood by 7.2% relative to the mean. This magnitude is 5.0% in Column 2, where we include additional individual-level controls such as worker age, education, and job role fixed effects, all measured as of February

2020.[13] The control variables all exhibit sensible signs. In particular, firms that had more remote work pre-COVID, smaller firms, and younger firms are more likely to see their employees spawning new firms post-COVID; as do firms that had higher spawning rates in 2019, a control we include to absorb unobserved employee spawning tendencies. In terms of employee characteristics, those with shorter job tenure, higher seniority, higher salary, prior founder experience, or younger and more educated employees are more likely to leave for entrepreneurship. These effects are consistent with determinants of entrepreneurial spawning documented in prior literature (e.g., Gompers et al. (2005), Babina et al. (2023), Babina and Howell (2024)).

Columns 3 and 4 show our firm-level cross-sectional results. We collapse both the dependent and independent variables from the individual level to the firm level. As such, the dependent variable is the share of employees (in percentage points) spawned between March 2020 and December 2022, and individual-level controls are now firm averages. Relative to the individual-level specification, the firm-level specification weighs each firm equally. We continue to find similar results with smaller magnitudes than individual-level results. A one-standard-deviation increase in RW_{post} increases firm-level spawning share by 3.7%.

Table 4 presents the instrumental variable results. Our primary instrument is firm-level average commute distance before COVID. The instrument is strong in the first stage, with a F-stat of 35 at the individual level and a F-stat of 299 at the firm level. Columns 1 and 3 of Table A.3 show the full first-stage results. A one-standard-deviation increase in $Commute$ increases RW_{post} by 4.8% (5.5%) standard deviations at the individual (firm) level. Larger, younger firms have higher RW during COVID, so are firms with higher pre-pandemic RW and spawning rates. Firms whose employees have shorter tenure, higher seniority, lower salary, or more past founder experience also tend to have higher RW during COVID. Table 4 shows the second-stage results. Based on Columns 2 and 4, a one-standard-deviation increase in RW_{post} increases worker-level likelihood of spawning by 0.28 p.p. and firm-level spawning share by 0.18 p.p., which correspond to 83% and 42% of their respective means. These effects are much larger than their OLS counterparts. We discuss magnitudes further below.

Table A.4 shows similar 2SLS results with our alternative instrument—county-level business closure orders during COVID. Note that this instrument requires us to drop county fixed effects, which could explain the larger estimates. The $BizClose$ instrument is strong at the firm level, with a F-stat of 87.1, and just reaching the conventional F-stat threshold of 10 at the individual level (F-stat=10.1). Columns 2 and 4 of Table A.3 present the full first-stage results for this instrument.

[13] Since age and education info is missing for 32% of individuals, we do not include these controls in our baseline specification.

Magnitude. Our estimated 2SLS effects are much larger than their OLS counterparts. However, their ratios are within the range surveyed by Jiang (2017) from the literature. The large 2SLS effects cannot be driven by weak instruments, as our first-stage F-stats are high. Instead, it could be explained by the presence of confounders that bias the OLS estimate downward relative to 2SLS. For example, companies with greater RW during COVID may also offer better unobserved non-wage amenities, hence better retaining their employees and reducing spawning. Another explanation is that our 2SLS estimates a local average treatment effect (LATE), which can be higher than the average treatment effect (ATE). This means that compliers have stronger responses to RW than the average firm. In our setting, compliers are firms that cater to their employees' commuting needs. Such firms may also monitor their employees less closely once remote (due to stronger employee bargaining power or a more employee-centered culture), leading to higher spawning sensitivity to remote work. Finally, the large 2SLS/OLS ratio could be explained by measurement error in RW, which generates attenuation bias in OLS but is corrected in 2SLS (Pancost and Schaller, 2022).

Our baseline 2SLS estimate of a 42% effect aligns with magnitudes found in other studies of entrepreneurship, which inherently has an outcome of low mean. For instance, Gottlieb et al. (2022) find a 45% increase in female entrepreneurship rate relative to the mean after the extension of job-protected maternity leave in Canada. Hombert et al. (2020) show that a French unemployment insurance reform increases the probability of an unemployed individual entering entrepreneurship by 58% relative to the mean. Also studying spawning, Babina and Howell (2024) show that a one-standard-deviation increase in corporate R&D increases employees' departure rate to entrepreneurship by 8.4 p.p., a 650% effect relative to the mean.

Instrument Validity. For our 2SLS to estimate a valid local average treatment effect (LATE), our instruments need to satisfy three conditions: 1) relevance, 2) monotonicity, and 3) exclusion.[14] We have shown that our instruments have a strong first-stage effect, with F-stats well above the conventional rule of thumb of 10.

The monotonicity condition implies that there are no "defiers" in our sample, i.e., no firms that adopt *lower* remote work during the pandemic in response to higher pre-pandemic employee commute distance. We test this following the methodology in Dobbie et al. (2018) (also used in González-Uribe and Reyes (2021) and Bias and Ljungqvist (2023)). The intuition is that monotonicity implies that the first-stage coefficient on the instrument should be non-negative in all subsamples formed based on observables. We test this implication in a variety of subsamples split by different firm characteristics. Table A.5 presents the result. We find that, across all subsamples, our in-

[14]Note that if the monotonicity assumption is violated, our 2SLS estimates would still be a weighted average of marginal treatment effects, but the weights would not sum to one (Angrist et al., 1996; Heckman and Vytlacil, 2005).

struments exhibit significantly positive first-stage coefficients, lending support to the monotonicity assumption.

The exclusion condition is ultimately untestable. Nevertheless, we provide some evidence consistent with this condition being likely to hold in our sample. Table A.6 shows that our two instruments have no significant correlation with firms' pre-pandemic spawning share or the share of employees with past founder experience. This suggests that, conditional on our controls, including RW in 2019, the instruments have no detectable correlations with unobservables that affect a firm's ex-ante spawning tendencies. This helps rule out the possibility that our instruments directly impact post-pandemic spawnings through their correlations with ex-ante unobservables, such as corporate culture, employee types, or telework infrastructure.[15] Note that even if there is a correlation that we failed to detect in Table A.6, such effects will still be controlled for as we control for pre-pandemic firm spawning rate and employee past founder experience in our analysis. Furthermore, our subsequent dynamic analysis shows that firms with different levels of *Commute* or *BizClose* trended similarly before the pandemic. Hence, any violation of the exclusion condition needs to be time-varying, and not be absorbed by the set of time-varying controls in our dynamic analysis (see Sections 3.2 and 4.2). Although we cannot rule out that such confounders could exist, the consistent results across our two instruments greatly reduce such remaining concerns. Finally, to the extent that the exclusion restriction is violated, our reduced form estimates can still be interpreted as the causal impact of commute distance (or local business closure orders) on post-pandemic spawning. These reduced form results are available in Table A.7.

4.2 Firm Panel Results

We next turn to dynamic evidence estimated from a firm-level panel. Table 5 shows the firm-level DID results estimated from Equations 3 and 5. We focus on a fixed set of individuals employed as of February 2020, and link their spawning activities over time to the RW policies of their Feb2020 employers, regardless of which employer they were with in a year. We estimate this on a balanced panel of Feb2020 firms from 2016 to 2022. Panel A includes firm and year fixed effects, and Panel B additionally includes industry-year fixed effects.

Columns 1-3 of Panel A show that individuals who experienced a greater RW increase with their Feb2020 employers were more likely to spawn into entrepreneurship post 2020. In particular, those whose Feb2020 employer had a one-standard-deviation higher ΔRW were 8.1% more likely to spawn post-2020 than pre-2020. This magnitude is 9.1% and 10.1% when we use *Commute* and *BizClose* as the treatment. Notably, because this sample fixes the set of individuals and tracks their spawning

[15] Our subsequent panel analysis also rules out endogeneity from unobserved employee types by holding employee composition constant.

over time regardless of their employers, the results are not driven by workers' selection into firms. In other words, our results cannot be explained by remote work increase attracting entrepreneurial workers who later became entrepreneurs (or driving away less-entrepreneurial workers).

Columns 4-6 of Panel A examine individuals' founder status, i.e., whether one stays in entrepreneurship after spawning. The specification follows Equation 5. The dependent variable is 100 times the fraction of individuals who were an entrepreneur in a year. Unlike spawning events, founder status is not mean reverting, and captures how long one stays in entrepreneurship in addition to the switch. We continue to find a positive effect. Those whose Feb2020 employer had a one-standard-deviation higher ΔRW were 12.5% more likely to be an entrepreneur post-2020 relative to the mean. This effect is 5.9% and 8.7% respectively when measuring treatment with $Commute$ and $BizClose$.[16] We find similar but smaller-magnitude results in Panel B of Table 5, which includes industry-year fixed effects.

Figure 4 shows the dynamics of the above DID results using ΔRW and $Commute$ as the treatment. We see a significant increase in spawning rate after the start of pandemic for firms with higher realized RW increase (Panel A), with the effect peaking in 2021. The effect is more persistent when using $Commute$ as the treatment (Panel C). Importantly, firms with different levels of RW increase or pre-pandemic commute distance trended similarly before the pandemic, lending support to the parallel trends assumption. Panels B and D examine the founder share, following the specification in Equation 6. We find a steadily increasing effect from 2020 to 2022. This reflects the cumulation of the annual spawning effects. It also suggests that the newly spawned individuals did not quickly return to wage employment; rather, they stayed in entrepreneurship, leading to a persistent effect of remote work on entrepreneurship. This points away from the marginally spawned businesses being transient and low-quality, a topic we further explore in Section 6. We observe similar patterns in Figure A.4 using $BizClose$ as the treatment.

4.3 Robustness

In this section, we investigate the robustness of our main results.

Including the largest firms. Our baseline analysis restricts to firms with 10 to 5000 employees as of February 2020. Panel A of Table A.8 shows the robustness of our results to including all firms above 10 employees.

Spawning before vs after departing wage job. About one-third of the spawning events in our sample happened before the worker formally left wage employment. This can occur either

[16]Table A.9 shows the "first-stage" results that correspond to our reduced-form panel DID. The panel restricts to 2019 to 2021 for which our RW measure is available.

because the employer permitted side entrepreneurship, or because the founder retroactively reported the new firm's start date after quitting and the startup emerged from "stealth" mode. If our results are all driven by side, part-time entrepreneurship while one holds a full-time job, it may call into question the quality of the spawned businesses, as well as whether there is any career risk associated with transitioning to entrepreneurship from wage employment. To examine this issue, we split our dependent variable based on whether the spawning occurred before or after the employee formally left her wage job. We then repeated our main analysis for these two outcomes in Table A.10. We find that the response is stronger for spawnings that happened after quitting than before quitting. Although experimentation with entrepreneurship could start well before a firm is formally launched, this finding alleviates the concern that remote work only drives side, part-time entrepreneurship.

Alternative RW measure. Our *RW* measure is based on traffic from residential, mobile, and VPN IPs relative to that from business IPs. One may be concerned that mobile and VPN traffic may misclassify some in-person work as remote—such as employees who travel for work or those using mobile networks in office. We note that mobile and VPN traffic is quite small relative to residential traffic in our data (Table A.11). Nevertheless, we show in Table A.12 that our results are similar using an alternative definition of *RW* based on residential and business IPs only (i.e., excluding mobile and VPN).

Within-firm variation across establishments. Our baseline IV analysis exploits variation in business closure orders in firms' headquarter counties. To tighten identification, we exploit the fact that multi-establishment firms may adopt different remote work policies for workers in different establishment locations, due to variation in local social distancing measures. This allows us to include firm fixed effects to absorb all potential firm-level confounders and exploit variation across establishments of the same firm. Because we do not have establishment-level *RW* measure, we conduct a reduced-form analysis exploiting variation in local business closure orders in *workers' locations*. We estimate the following individual-level specification on workers with known Feb2020 locations:

$$Spawn_{i,f,c} = \alpha_f + \gamma Worker's\ Local\ BizClose_{i,c} + \lambda \boldsymbol{X_i} + \epsilon_i \tag{7}$$

In this equation, c represents the worker's county in Feb2020 and α_f represents firm fixed effects for the Feb2020 employer, which absorbs all firm-level controls (i.e., $\boldsymbol{X_f}$ in Equation 1). $Worker's\ Local\ BizClose_{i,c}$ is business closure orders in the worker's county. Table A.13 reports the results.[17] We find that, within the same firm, workers located in counties with stricter social distancing measures were more likely to spawn into entrepreneurship subsequently. Importantly, any firm-level unobservables, such as culture, employee types, or other firm policies cannot drive this finding. This within-firm analysis also addresses any concerns about measurement errors in our

[17] The sample size is smaller because not all individuals report a location on LinkedIn.

firm-level RW measure.

Fixing employees in 2016. Our firm-level dynamic analysis is based on a fixed set of employees employed in February 2020. One potential concern with this sample is that these individuals may have limited spawning activities between 2016 to 2019, given that they were wage employed in early 2020. This may limit our ability to detect any significant pre-trend. One way to address this is to fix employees in 2016 instead of 2020. Specifically, we can focus on Feb2020 firms' employees back in January 2016. These employees should face less constraint in spawning in 2016–2019 than the Feb2020 employees. However, a downside of this approach is that many of these 2016 employees may have already left the Feb2020 firm by 2020, making our RW measure less relevant for their post-2020 spawning decisions. In other words, the "intent-to-treat" on this sample is noisier. Nevertheless, we show in Figure A.5 that similar results continue to hold in this sample, though, as expected, the post-2020 treatment effects are noisier.

4.4 Alternative Explanations

We also consider and rule out a few alternative explanations.

COVID-specific demand shocks. One potential explanation of our results is that they reflect COVID-specific demand shocks rather than workers' labor supply choices. For example, employees at firms more amenable to remote work may possess skills that became more valuable during the pandemic (e.g., telework/communications or food delivery sectors), changing the returns to entrepreneurship. Addressing this concern, our cross-sectional analysis compares workers working in the same industry and with the same job role; our DID analysis includes the interactions between firm characteristics and the post-2020 dummy, allowing COVID shocks to vary by firm types. Industry-year fixed effects in Panel B of Table 5 further address this concern by absorbing unobserved industry-specific shocks.

To further make sure that our results do not reflect COVID-specific booms, we show that our results are robust to dropping industries that benefited the most from COVID. Specifically, using Census BDS data, we identify the top 60 NAICS 4-digit industries (representing 20% of all NAICS 4-digits) that had the highest job creations in 2020-2022 relative to 2018-2019. Panel C of Table A.8 shows that our results are robust to dropping firms in these top booming industries. Finally, our analysis conditional on turnovers in Section 5 also helps to rule out unobserved demand shocks, as these shocks should lead to labor reallocations in general, not just entry into entrepreneurship.

Remotability of spawned firms. Another alternative explanation is that our results are driven by the remotability of the spawned firm, which may correlate with remote work of the employer if spawning happens in the same industry or location. If this is the case, remote work opportunities may directly impact the returns to entrepreneurship, regardless of the ability to spawn. Our analysis compares workers working in the same industry or with the same job role. In Table A.14, we further include fixed effects for workers' locations (counties). This specification compares workers living in the same location but working for firms with different remote work policies. As such, local opportunities to start a remote business should not drive our findings. We continue to find similar results. Finally, we show in Section 6 that the marginally spawned firm is just as likely to be remote as in-person.

5 Is it Unique to Entrepreneurship?

One may argue that some of the forces that drive the effect of remote work on entrepreneurship may also drive worker turnovers in general, including turnover into other wage jobs or unemployment. To assess the extent to which our results are unique to entrepreneurship, we condition our individual-level analysis on those experiencing job turnovers and examine whether remote work *disproportionately* shifts individuals into entrepreneurship relative to other labor destinations. Specifically, we restrict to individuals who left their Feb2020 employer between March 2020 and December 2022 (about 16% of our individual-level sample). We then repeat our individual-level cross-sectional analysis on this subsample. This conditional analysis ensures that we are not just capturing a general turnover effect; rather, any effect reflects mechanisms unique to entrepreneurship.

This conditional analysis has two additional benefits. First, conditioning on a turnover, i.e., the individual updating her resume on LinkedIn, addresses any concerns about truncation issues from stale resumes in the LinkedIn data. Second, it also helps to rule out that our results are driven by COVID-specific demand for employees of certain skills that correlate with their employer's remote work policy. Such demand-side factors should affect not just entry into entrepreneurship but also the general reallocation of workers across sectors and firms, which our conditional sample captures. As such, any demand-side factors need to be specific to entrepreneurship to explain our conditional results.

Table 6 presents the 2SLS results of this conditional analysis (OLS results are in Table A.15). We find that, conditional on leaving their Feb2020 employer between March 2020 and December 2022, workers initially more exposed to remote work were more likely to pursue entrepreneurship relative to being unemployed or wage employed with another firm. Based on Column 2, a one-standard-deviation increase in RW_{post} increases departures to entrepreneurship relative to other destinations by 30% relative to the mean. This finding suggests that the mechanisms through which remote work spurs entrepreneurship is somewhat unique to the economics of entrepreneurship. We explore this more next.

6 Potential Mechanisms

In this section, we investigate the mechanisms through which remote work spurs entrepreneurial spawning. We identify three possible mechanisms: 1) preference change 2) experimentation, and 3) forced entrepreneurship. Overall our evidence points toward the dominant role of the experimentation channel, where remote work gives employees downside protection and, possibly, the slack time needed for entrepreneurial experimentation.

Preference change. One channel for our results is that remote work increases workers' preference for flexibility or a "quiet life," which entrepreneurship may offer.[18] This, however, should only predict spawning into low-growth, hobby-based self-employment that offers substantial non-pecuniary benefits (Schoar, 2010). In contrast, high-growth, transformational entrepreneurship is time-consuming and requires founders' full commitment.

We directly examine the quality of the marginally spawned firms. We employ two quality measures: initial employment and whether the business received subsequent VC backing.[19] In our sample, the average spawned firm has 0.9 initial employees excluding the founders (note that LinkedIn may not capture all employees), and 6% of them subsequently received VC funding—much higher than the VC backing rate among the average new firms in the economy (<1%). Our first approach conditions on spawned individuals and examines how remote work changes the average quality of the spawned firms. Our second approach splits the spawning outcome by whether the spawned firm is of high or low quality. We then compare these two types of spawning responses to test if the *marginally* spawned firm tends to be of high quality. If the preference channel drives our results, we should observe remote work decreasing the average quality of the spawned firms, and a much stronger response in low-quality spawning than high-quality spawning.

Table 7 presents the result based on 2SLS (OLS results in Table A.16). Panel A conditions on spawned individuals and shows that remote work increases the average quality of spawned firms. In particular, a one-standard-deviation increase in RW_{post} increases the initial employment of spawned firms by 38%, and increases the likelihood of receiving future VC funding by 171%. This shift in average entry quality, however, could reflect either a decrease in low-quality spawning or an increase in high-quality spawning. Panel B distinguishes the two.

In Panel B, we split the spawning outcome by our quality measures. We compare the percentage effects of a one-standard-deviation increase in RW_{post} on each spawning type relative to the

[18]Alternatively, flexibility offered by remote work could substitute flexibility offered by self-employment, thereby *reducing* flexibility-based entrepreneurship (Barrios et al., 2024).

[19]We define initial employment as the maximum employment within the initial two years of founding. We identify VC-backed as firms that can be matched to the VC-backed universe in Pitchbook as of 2024.

outcome mean. If remote work does not change the average quality of spawned firms, we should observe the same percentage effects on high- and low-quality spawnings. Panel B shows that remote work significantly increases both high- and low-quality spawning, but the percentage effect is much stronger for high-quality spawning, consistent with the finding in Panel A. A one-standard-deviation increase in RW_{post} increases spawning into employer firms by 184% while increasing spawning into non-employer firms by 96% (Columns 1-2). The spawning response is four times larger for VC-backed firms than non-VC backed firms (Columns 3-4). These differences are also statistically significant, as indicated by the P-values. These results suggest that remote work does not primarily spawn low-quality businesses, which tend to be associated with greater flexibility or other nonpecuniary benefits. Rather, the marginally spawned firm is of higher quality than the average new firm. These results also corroborate our finding in Figure 4 that the spawned entrepreneurship is not transient.

One concern is that LinkedIn may not capture all entrepreneurial activity in two ways: first, entrepreneurs who explored ideas but never formally registered a business (pre-launch failures), and second, businesses that launched but quickly failed. For pre-launch failures, this limitation aligns with our research focus on actual business formation rather than early-stage exploration. Regarding quick failures, we acknowledge that LinkedIn data might not fully capture the left tail of the quality distribution. However, this potential left-tail truncation is unlikely to bias the quality comparisons in Table 7, as all groups and outcomes analyzed are observed within the same potentially truncated LinkedIn sample. As such, our comparative analysis remains valid despite LinkedIn's potential underrepresentation of short-lived businesses.

Another possible preference story is that workers, after prolonged exposure to remote work, miss in-person interactions and start their own businesses to fulfill their social needs. This should predict that the marginally spawned business is more likely to be in-person than remote. To test this, we split the spawning outcome by whether the spawned firm had an above- or below-median RW within 2 years of spawning. Due to their nascency, we do not observe RW for many spawned firms. Hence, we do this split only within the set of spawned firms for which we can reliably measure RW.[20] Table A.17 shows the results. We find that, across both OLS and 2SLS, the spawning response is similar for high-RW and low-RW new firms, with no statistically significant difference. As such, a preference for in-person interactions is unlikely to explain our results.

Taken together, the evidence suggests that preference changes are unlikely to explain the positive effect of remote work on entrepreneurial spawning. We next explore non-preference-based channels.

[20]Presumably, the ones with missing RW are likely to be firms with few or no employees. These firms should go against the story that entrepreneurs found them to socialize with co-workers.

Experimentation. Remote work could provide workers the time and "stealth" needed to experiment with entrepreneurship without risking their current careers. First, remote work frees up time by saving on commuting time and increasing productivity, which reduces actual work hours.[21] Remote work also frees up time by increasing flexibility: employees can work on their side projects during lunch breaks or lulls in their regular duties. This slack time gives workers the opportunity to develop and tinker with entrepreneurial ideas (Agrawal et al., 2018). Remote work can also provide downside protection for experimentation. Remote work provides employees with more privacy and reduced employer monitoring, lowering the risk that a side project will be discovered by their employer, which could negatively impact their career.[22] These factors combined allow employees to more effectively maintain their current job as a fallback option while exploring entrepreneurship—an arrangement that was much less feasible when working in an office environment. From this perspective, this channel is similar to the career risk channel in Gottlieb et al. (2022), where job-protected leave increases workers' experimentation with entrepreneurship by providing downside protection. Here, we can think of remote work as a flexible form of "job-protected leave."

If the downside protection channel is at work, our result should be stronger in industries with a higher risk of failure, as the option value to experiment before committing is higher in these industries. Table 8 explores such heterogeneity, using the same methods as Table 7. We measure industry risk with annual probabilities of failure by young firms (age$\leq$5) in a NAICS 4-digit industry, obtained from the Business Dynamic Statistics (BDS) from U.S. Census. Panel A of Table 8 conditions on entry and shows that remote work shifts entry towards riskier industries. Panel B partitions spawning by industry risk at the median, and compares the spawning responses into high- vs low-risk industries. We find a much stronger spawning response into riskier industries than into safer industries relative to their respective means (143% vs 64%), with the difference being statistically significant. We find similar OLS results in Table A.16. These results suggest that remote work disproportionately enables experimentation in riskier industries, where downside protection from the wage job is more valuable.[23]

If remote work enables experimentation by providing more time and flexibility, our results should be stronger for entrepreneurship that requires greater time commitment, such as innovative, high-growth startups. More flexible and less time-consuming entrepreneurial activities, such as selling crafts on Etsy or running an Airbnb, should respond less to remote work since these ventures

[21] Barrero et al. (2023) report that the average daily savings in commuting and grooming time is 65 minutes for American workers. The literature typically finds that hybrid arrangements (working from home some days of the week) increase worker productivity, while fully remote arrangements do so to a lesser extent, though the lower productivity is often offset by savings in commute time. See detailed review by Barrero et al. (2023).

[22] Several studies (Gibbs et al., 2023; Yang et al., 2022; Parker, 2023; Emanuel et al., 2023) find that remote work leads to fewer contacts and less communication within the organization and reduces mentoring.

[23] Consistent with experimentation, we also find substantial entry into industries outside the employer industry, as shown in Table A.18.

could be managed even while working an in-person job. Our prior finding that the marginally spawned firm is more likely to be high-quality, supports the idea that remote work relaxes time and flexibility constraints for potential entrepreneurs.

To provide further direct evidence on the slack time channel, we exploit variation in local schools' learning model during the pandemic to generate variation in slack time for working parents. Parents whose children attended virtual or hybrid schooling for longer periods during the pandemic would have less time for entrepreneurial experimentation while working from home. We obtain data on K12 schools' learning model during the pandemic from `www.covidschooldatahub.com` (Jack et al., 2023), which provides the fraction of time each school operated in-person, hybrid, or virtually during 2020-2021, along with each school's location. We collapse these data to the county level to obtain the fraction of time local schools were in person and use it as an interaction variable in our baseline individual-level analysis. Table A.19 shows the results. We find that workers of age 30-45—those most likely to have school-age children—had a stronger spawning response if their local K12 schools were in-person more during the pandemic. This is consistent with in-person schooling freeing up more time for parents working remotely. In contrast, we do not find such an interaction effect for workers outside of the 30-45 age range, who were unlikely to have school-age children at home.[24]

Forced entrepreneurship. One last possibility is that remote work induces forced entrepreneurship by triggering layoffs or involuntary turnovers. For example, firms that increased remote work more may also have laid off more workers, who in turn started a business out of necessity. This explanation is unlikely to be accurate given that, during COVID, firms that pivoted more effectively to remote work adapted better, while those relying heavily on in-person work suffered more and experienced more layoffs (Forsythe et al., 2020; Mongey et al., 2021).

Another possibility is that high-RW firms experienced more quitting as they tried to bring workers back to the office in 2022 (Barrero et al., 2021a), and these quitting workers later started a business. To the extent that this quitting is driven by a preference for flexibility or "quiet life," it amounts to the preference channel, which we ruled out above. Note that this explanation also implies that spawning should occur with a substantial delay after firms began reverting to in-person operations.

To further test the forced entrepreneurship channel, we restrict our main analysis to firms that experienced continued yearly employment growth from 2020 to 2022, i.e., firms that were unlikely to have had mass layoffs during COVID. Panel B of Table A.8 reports the results. We find effects that are larger, rather than smaller, than those in our main sample. This suggests that our main

[24]We do not present the IV results as our instruments are not strong enough for the interaction term.

results are unlikely to be driven by forced entrepreneurship.

Taken together, the evidence in this section suggests that remote work spawns entrepreneurship mainly by providing the time and downside protection needed for entrepreneurial experimentation. Employee selection, preference change, or forced turnovers are unlikely to explain our results.

While we find a positive overall effect of remote work on entrepreneurial spawning, we acknowledge that remote work could potentially have opposing effects. For example, Barrios et al. (2024) propose a substitution effect, whereby the flexibility of remote work in traditional employment reduces entrepreneurial entry by offering similar nonpecuniary benefits without the associated risks. In our sample, the marginal entrepreneurship is unlikely to be flexibility-based. As such, any negative substitution effect appears to be outweighed by countervailing positive mechanisms, leading to the net positive effect we observe.

7 Aggregate Effects

7.1 Industry/County-Level Analysis

We validate our micro-level evidence with aggregate-level evidence based on US Census data and industry-level job remotability. The advantage of this analysis is that we can address any potential concerns about LinkedIn not capturing all workers or all new businesses, or capturing them with time truncation; we can also address any concerns about measurement errors with our Internet-based RW measure. This analysis also helps verify whether our micro-level evidence can aggregate to the macro level.[25] The downside, however, is that we have to make the assumption that spawning tends to happen in the same industry as the prior employer, since we only observe the industry of the new firm, not of the entrepreneur's prior employer. In our data, 18% (23%) of spawned entrepreneurs are in the same NAICS 3-digit (2-digit) as their previous employer. This is high given that there are 102 NAICS 3-digits and 20 NAICS 2-digits, implying a same-industry probability of only 0.98% and 5% if spawning is random.

Industry-Level Firm Entry. We first examine how changes in industry-level new firm entry around COVID vary with an industry's remotability—the extent to which its jobs can be performed at home or remotely (Dingel and Neiman, 2020). We estimate a dynamic DID of the following

[25] For example, if most of the firm-level variation in RW is within industry (or county) rather than between them, then a shift to remote work wouldn't generate any sectoral or regional differences in entrepreneurship rates.

specification at the NAICS-3digit-year level from 2016 to 2022:

$$Ln(no.\ of\ new\ employer\ firms)_{n,t} = \alpha_n + \beta_t + \sum_{t \neq 2019}^{2016 \to 2022} \theta_t \times Remotability_n \times \mathbb{1}(Year = t) + \epsilon_{n,t} \quad (8)$$

, where α_n indicates industry fixed effects and β_t indicates year fixed effects. The dependent variable is the log number of new employer firms started in a NAICS-3digit-year based on US Business Dynamic Statistics (BDS). *Remotability* comes from Dingel and Neiman (2020) and is the average remotability of an industry's jobs. It is standardized before interacting with year indicators.

Figure 5 presents the results. We find that more remotable industries experienced higher new firm entry from 2020 to 2022 relative to pre-2020, yet they trended similarly as less remotable industries before 2020. The interpreting assumption is that workers tend to stay in the same industry when spawning; to the extent they do not, the measurement error should bias against finding an effect. Of course, one interpretation is that more remotable industries are more desirable during COVID, hence experiencing greater new firm entry. To the extent this desirability is preference-driven, we already ruled it out in Section 6. We also showed in Table A.17 that spawning does not flow predominantly into industries with high levels of remote work.

County-Industry Level Employment by New Firms. To further zoom in on quality-adjusted new firm entry, we also examine employment at new firms (age 0-1) from the US Quarterly Workforce Indicators (QWI).[26] This measure, essentially employment-weighted new firm count, ensures that any result we find does not reflect low-quality entry. Additionally, QWI data allow us to conduct this analysis at the industry-county level, which allows for richer fixed effects to absorb potential confounders.

We estimate a dynamic DID of the following specification at the industry(NAICS 2digit)-county-year level:

$$Ln(employment\ at\ new\ firms)_{n,c,t} = \alpha_{n,c} + \beta_{c,t} + \sum_{n \neq 2019}^{2015 \to 2023} \theta_t \times Remotability_n \times \mathbb{1}(Year = t) + \epsilon_{n,c,t} \quad (9)$$

, where $\alpha_{n,c}$ indicates industry-county fixed effects and $\beta_{c,t}$ indicates county-year fixed effects. Similar to Equation 8, we use industry-level *Remotability* from Dingel and Neiman (2020) as the treatment variable. The dependent variable is log total employment by new firms started in a NAICS-2-digit-county-year based on QWI data. The sample is from 2015 to 2023, the last available year of QWI. Figure A.6 plots the dynamic DID results. We find that total new firm employment

[26]We find similar results expanding to the employment at firms of age 0-3.

increased more in more remotable industries. This result confirms our BDS result using quality-adjusted entry.

Note that our aggregate-level results do not contradict Barrios et al. (2024), who find decreased business applications post-2020 in zipcodes with higher remotability of jobs. Based on Business Formation Statistics from US Census, only one-third of the business applications during the pandemic were by likely employers, and only 11.9% were by incorporated businesses. In other words, the majority of business applications during COVID never materialized into economically meaningful firms with paid employees. Given that non-employer, unincorporated businesses (e.g., sole proprietorship) offer much higher flexibility than incorporated employer firms, this explains why flexibility from remote work can substitute flexibility-based entrepreneurship in their sample. Such a substitution effect is unlikely to dominate in our aggregate sample because we focus on employer firms and quality-adjusted entry. Our micro-level results in Section 6 also ruled out that the preference for flexibility drives our results.

7.2 Calibrate to the Macro Time Series

How much of the post-COVID increase in startup rate could be explained by the shift to remote work? We conduct a back-of-the-envelope calculation to evaluate this. Based on Figure 1, the average RW across firms increased by about 0.123 from pre-COVID to post-COVID. This translates to a 7.6% increase in spawning rate based on our cross-sectional estimate in Column 1 of Table 3. There were about 130 million full-time employees in the US before COVID. Based on Census Business Formation Statistics, the average annual number of new business applications increased from 3.51 million pre-COVID to 4.96 million post-COVID. Based on Current Population Survey (CPS) data, 63% of entrepreneurs come from wage employment pre-COVID.[27] Thus, the implied annual spawning rate is 1.7% (=3.51*0.63/130). A 7.6% RW-induced increase in this spawning rate would imply 167.8K (=1.7%*7.6%*130000) additional business applications per year. Given that the actual increase in annual business applications is 1.45mil from pre-COVID to post-COVID, our RW-based estimate can account for 11.6% of this increase.

7.3 Allocative Implications

We cannot directly comment on the distributional effects of remote work-induced spawning. On one hand, spawnings from remote work could simply reflect a reallocation of innovative activities

[27]Based on CPS data from 2016 to 2019, the period before COVID, we compute the fraction of entrepreneurs each year who were in wage employment in the previous year. This number is consistent with the estimate from the Kauffman Foundation, which is between 60% and 70%.

that would have happened in-house, in which case there will be no changes in aggregate output. On the other hand, it is possible that the spawned entrepreneurs pursue innovations that would not have happened inside the prior employer, due to frictions in innovation incentives (Gompers et al., 2005; Seru, 2014). Indeed, we show in Table A.18 that there is a significant spawning effect even outside the industry of the original employer. It is also possible that remote work relaxes constraints on employees' ability to explore their outside options in entrepreneurship, leading to better matches. In these cases, spawnings could reflect more efficient reallocations. Regardless, the impact on incumbent firms is likely to be negative if spawning destroys firm-specific knowledge, and/or if firms cannot adequately capture the surplus released by spawning (Ma et al., 2023). As such, firms may underinvest in remote work relative to the social optimum. We leave such allocative effects to future research.

8 Conclusion

The majority of entrepreneurs are prior wage workers. How work is organized in wage employment, therefore, impacts workers' decision to become an entrepreneur. This paper shows that the recent paradigm shift to remote work increased wage workers' transitions to entrepreneurship. Using big data on Internet activities, we create a novel measure of firm-level remote work. We then link it to LinkedIn data to test how variations in firms' remote work policies affect workers' transition to entrepreneurship, i.e., entrepreneurial spawning. We find that firms with higher levels of remote work during the pandemic saw a higher share of their employees starting new firms. Marginally created new firms tend to be of higher quality than the average new firm. The spawning effect of remote work also holds conditional on job turnovers, suggesting remote work directs workers disproportionately to entrepreneurship relative to other labor outcomes. These effects are not driven by employee selection, preference change, or forced turnover. Rather, remote work provides the time and downside protection needed for entrepreneurial experimentation, allowing workers to better use their wage job as a fallback option when exploring entrepreneurship. We estimate that at least 11.6% of the post-pandemic increase in new firm entry can be explained by the rise of remote work. Firms and policymakers need to take such spillover effects into account when designing future work policies. Our findings have important implications for the talent flows between wage employment and entrepreneurship in a new labor market equilibrium shaped by remote work.

References

Agrawal, A., C. Catalini, A. Goldfarb, and H. Luo (2018). Slack time and innovation. *Organization Science 29*(6), 1056–1073.

Agrawal, A., I. Hacamo, and Z. Hu (2021). Information dispersion across employees and stock returns. *The Review of Financial Studies 34*(10), 4785–4831.

Aksoy, C. G., J. M. Barrero, N. Bloom, S. J. Davis, M. Dolls, and P. Zarate (2022). Working from home around the world. *Brookings Papers on Economic Activity 2022*(2), 281–360.

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996). Identification of causal effects using instrumental variables. *Journal of the American statistical Association 91*(434), 444–455.

Atkin, D., A. Schoar, and S. Shinde (2023). Working from home, worker sorting and development. Technical report, National Bureau of Economic Research.

Babina, T. (2020). Destructive creation at work: How financial distress spurs entrepreneurship. *Review of Financial Studies 33*(9), 4061–4101.

Babina, T. and S. T. Howell (2024). Entrepreneurial spillovers from corporate r&d. *Journal of Labor Economics 42*(2), 469–509.

Babina, T., P. Ouimet, and R. Zarutskie (2023). IPOs, human capital, and labor reallocation. *Journal of Financial and Quantitative Analysis*.

Barrero, J. M., N. Bloom, and S. J. Davis (2021a). Let me work from home, or i will find another job. *University of Chicago, Becker Friedman Institute for Economics Working Paper* (2021-87).

Barrero, J. M., N. Bloom, and S. J. Davis (2021b). Why working from home will stick. Technical report, National Bureau of Economic Research.

Barrero, J. M., N. Bloom, and S. J. Davis (2023). The evolution of work from home. *Journal of Economic Perspectives 37*(4), 23–49.

Barrios, J. M., Y. Hochberg, and H. L. Yi (2024). Hustling From Home? Work From Home Flexibility and Entrepreneurial Entry. Technical report, National Bureau of Economic Research.

Barrios, J. M., Y. V. Hochberg, and H. Yi (2022). Launching with a parachute: The gig economy and new business formation. *Journal of Financial Economics 144*(1), 22–43.

Bernstein, S., R. R. Townsend, and T. Xu (2024). Flight to safety: How economic downturns affect talent flows to startups. *The Review of Financial Studies 37*(3), 837–881.

Bias, D. and A. Ljungqvist (2023). Great recession babies: How are startups shaped by macro conditions at birth? *Swedish House of Finance Research Paper* (23-01).

Bloom, N., R. Han, and J. Liang (2024). Hybrid working from home improves retention without damaging performance. *Nature*, 1–6.

Bloom, N., J. Liang, J. Roberts, and Z. J. Ying (2015). Does working from home work? Evidence from a Chinese experiment. *The Quarterly journal of economics 130*(1), 165–218.

Brucks, M. S. and J. Levav (2022). Virtual communication curbs creative idea generation. *Nature 605*(7908), 108–112.

Brueckner, J. K., M. E. Kahn, and G. C. Lin (2023). A new spatial hedonic equilibrium in the emerging work-from-home economy? *American Economic Journal: Applied Economics 15*(2), 285–319.

Buckman, S. R., J. M. Barrero, N. Bloom, and S. J. Davis (2025). Measuring work from home. Technical report, National Bureau of Economic Research.

Chen, A., M. B. Zhang, and Z. Zhang (2023). Talent market competition and firm growth. *Available at SSRN*.

Chen, C., C. B. Frey, and G. Presidente (2022). Disrupting science. Technical report.

Decker, R. and J. Haltiwanger (2023). Surging business formation in the pandemic: Causes and consequences. *Brookings Papers on Economic Activity*, 3–24.

Decker, R., J. Haltiwanger, R. Jarmin, and J. Miranda (2014). The role of entrepreneurship in us job creation and economic dynamism. *Journal of Economic Perspectives 28*(3), 3–24.

Dingel, J. I. and B. Neiman (2020). How many jobs can be done at home? *Journal of Public Economics 189*, 104235.

Dobbie, W., J. Goldin, and C. S. Yang (2018). The effects of pre-trial detention on conviction, future crime, and employment: Evidence from randomly assigned judges. *American Economic Review 108*(2), 201–240.

Duchin, R. and D. Sosyura (2021). Remotely productive: The efficacy of remote work for executives. *Available at SSRN 3761972*.

Eisfeldt, A. L., G. Schubert, M. B. Zhang, and B. Taska (2023). The labor impact of generative ai on firm values. *Available at SSRN 4436627*.

Emanuel, N. and E. Harrington (2023). Working remotely? selection, treatment, and the market for remote work. *Selection, Treatment, and the Market for Remote Work (May 1, 2023). FRB of New York Staff Report* (1061).

Emanuel, N., E. Harrington, and A. Pallais (2023). The power of proximity to coworkers: training for tomorrow or productivity today? Technical report, National Bureau of Economic Research.

Feldman, M. P. (2013). *The Geography of Innovation*, Volume 2. Springer Science & Business Media.

Florida, R. and C. Mellander (2016). Rise of the startup city: The changing geography of the venture capital financed innovation. *California Management Review 59*(1), 14–38.

Flynn, S., A. C. Ghent, and V. Nair (2024). Determinants and consequences of return to office policies. *Available at SSRN*.

Forsythe, E., L. B. Kahn, F. Lange, and D. Wiczer (2020). Labor demand in the time of covid-19: Evidence from vacancy postings and ui claims. *Journal of public economics 189*, 104238.

Gibbs, M., F. Mengel, and C. Siemroth (2023). Work from home and productivity: Evidence from personnel and analytics data on information technology professionals. *Journal of Political Economy Microeconomics 1*(1), 7–41.

Glaeser, E. L. and N. Hausman (2020). The spatial mismatch between innovation and joblessness. *Innovation Policy and the Economy 20*(1), 233–299.

Gompers, P., J. Lerner, and D. Scharfstein (2005). Entrepreneurial spawning: Public corporations and the genesis of new ventures, 1986 to 1999. *Journal of Finance 60*(2), 577–614.

González-Uribe, J. and S. Reyes (2021). Identifying and boosting âgazellesâ: Evidence from business accelerators. *Journal of Financial Economics 139*(1), 260–287.

Gottlieb, J. D., R. R. Townsend, and T. Xu (2022). Does career risk deter potential entrepreneurs? *The Review of Financial Studies 35*(9), 3973–4015.

Hacamo, I. and K. Kleiner (2022). Forced entrepreneurs. *The Journal of Finance 77*(1), 49–83.

Hansen, S., P. J. Lambert, N. Bloom, S. J. Davis, R. Sadun, and B. Taska (2023). Remote work across jobs, companies, and space. Technical report, National Bureau of Economic Research.

Heckman, J. J. and E. Vytlacil (2005). Structural equations, treatment effects, and econometric policy evaluation 1. *Econometrica 73*(3), 669–738.

Hombert, J., A. Schoar, D. Sraer, and D. Thesmar (2020). Can unemployment insurance spur entrepreneurial activity? evidence from france. *The Journal of Finance 75*(3), 1247–1285.

Jack, R., C. Halloran, J. Okun, and E. Oster (2023). Pandemic schooling mode and student test scores: evidence from us school districts. *American Economic Review: Insights 5*(2), 173–190.

Jeffers, J. S. (2024). The impact of restricting labor mobility on corporate investment and entrepreneurship. *The Review of Financial Studies 37*(1), 1–44.

Jiang, W. (2017). Have instrumental variables brought us closer to the truth. *Review of Corporate Finance Studies 6*(2), 127–140.

Kerr, W. R., R. Nanda, and M. Rhodes-Kropf (2014). Entrepreneurship as experimentation. *Journal of Economic Perspectives 28*(3), 25–48.

King, R. G. and R. Levine (1993). Finance and growth: Schumpeter might be right. *Quarterly Journal of Economics 108*(3), 717–737.

Kwan, A., B. Matthies, and A. Yuskavage (2023). Measuring the impact of remote work using big data. *working paper*.

Lin, Y., C. B. Frey, and L. Wu (2023). Remote collaboration fuses fewer breakthrough ideas. *Nature 623*(7989), 987–991.

Ma, S., W. Wang, and Y. Wu (2023). Steering labor mobility through innovation. Technical report.

Mi, X., X. Feng, X. Liao, B. Liu, X. Wang, F. Qian, Z. Li, S. Alrwais, L. Sun, and Y. Liu (2019). Resident evil: Understanding residential ip proxy as a dark service.

Mongey, S., L. Pilossoph, and A. Weinberg (2021). Which workers bear the burden of social distancing? *The Journal of Economic Inequality 19*(3), 509–526.

Pancost, N. A. and G. Schaller (2022). Measuring measurement error. *Available at SSRN 4045772*.

Parker, K. (2023). About a third of us workers who can work from home now do so all the time.

Ramani, A. and N. Bloom (2021). The donut effect of covid-19 on cities. Technical report, National Bureau of Economic Research.

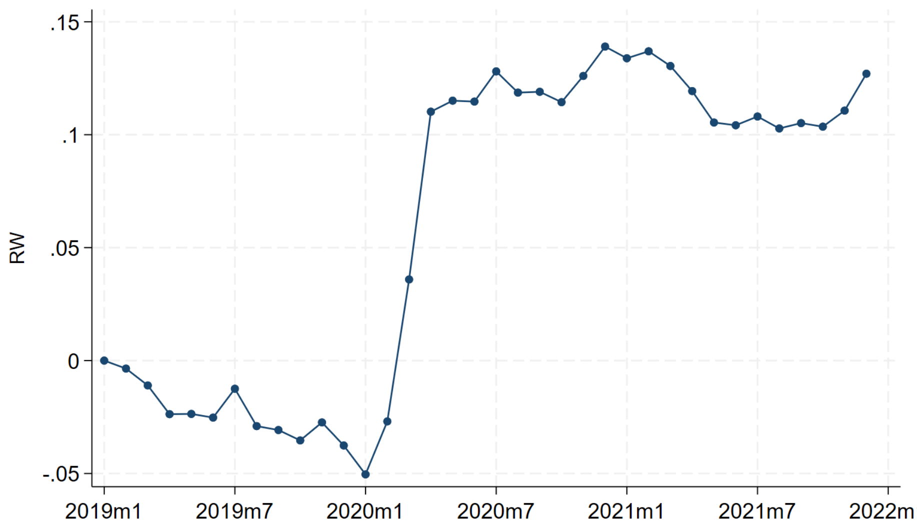
Schoar, A. (2010). The divide between subsistence and transformational entrepreneurship. *Innovation policy and the economy 10*(1), 57–81.

Seru, A. (2014). Firm boundaries matter: Evidence from conglomerates and r&d activity. *Journal of financial economics 111*(2), 381–405.

Spiegel, M. and H. Tookes (2021). Business restrictions and covid-19 fatalities. *The Review of Financial Studies 34*(11), 5266–5308.

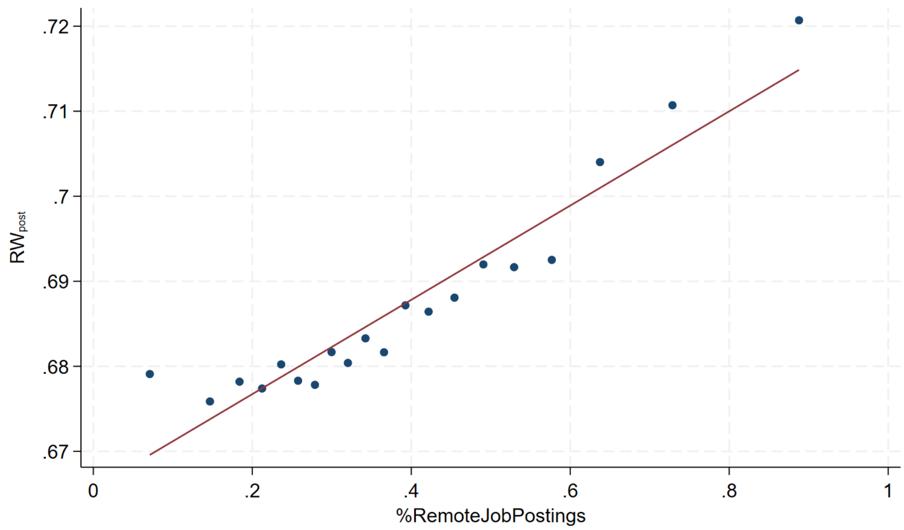
Yang, L., D. Holtz, S. Jaffe, S. Suri, S. Sinha, J. Weston, C. Joyce, N. Shah, K. Sherman, B. Hecht, et al. (2022). The effects of remote work on collaboration among information workers. *Nature human behaviour 6*(1), 43–54.

Figure 1: Changes in Average Remote Work from 2019 to 2021



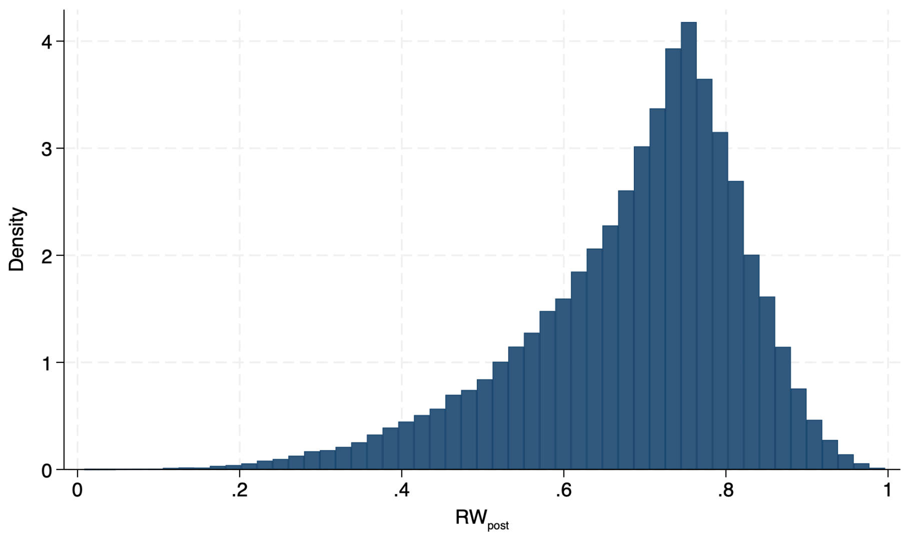
This graph shows the evolvement of average firm-level RW measure from January 2019 to December 2021, with January 2019 normalized to 0. We compute the monthly averages among firms active in February 2020.

Figure 2: RW_{post} and Firm-Level Fraction of Remote Job Postings



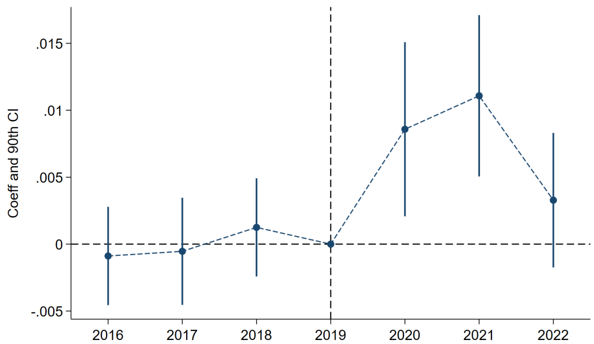
This graph shows the binned scatter plot between our firm-level RW_{post} measure and firm-level fraction of job postings mentioning remote work over 2020-2021. Appendix A.3 details how we identify remote work arrangement from job posting data using LLM, following the method in Hansen et al. (2023). RW_{post} is a firm's average RW in 2020 and 2021. The binned scatter plot controls for RW_{pre}, firm's average RW in 2019. Firm-level data is binned into 20 quantiles based on the value of $\%RemoteJobPostings$. The height of each dot represents the average RW_{post} within the bin.

Figure 3: Histogram of $\mathrm{RW}_{\mathrm{post}}$

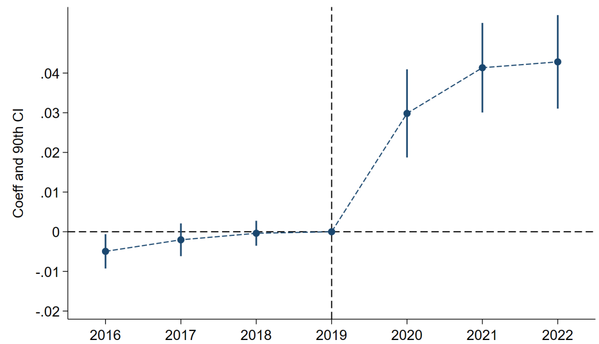


This graph shows the histogram of firm-level average remote work in 2020 and 2021 (RW_{post}) for firms in our main sample.

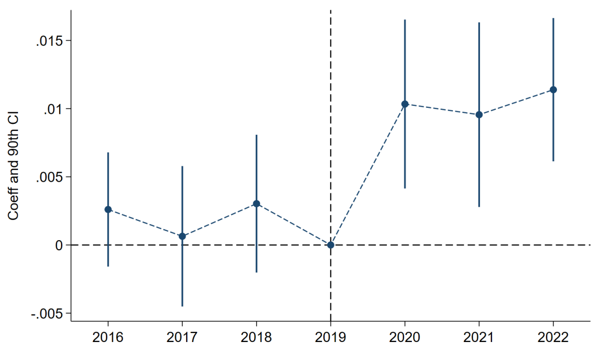
Figure 4: Firm-Level DID Dynamics



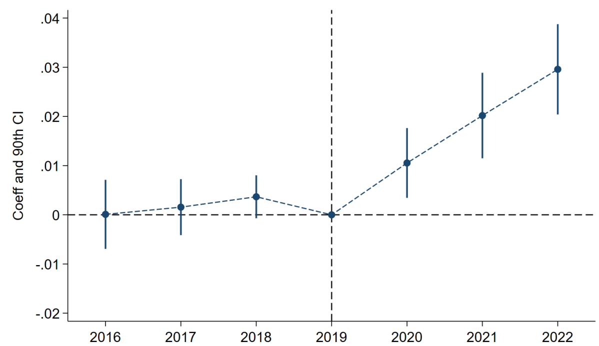
(a) Treat $= \Delta RW$, Outcome $= SpawnShare$



(b) Treat $= \Delta RW$, Outcome $= FounderShare$



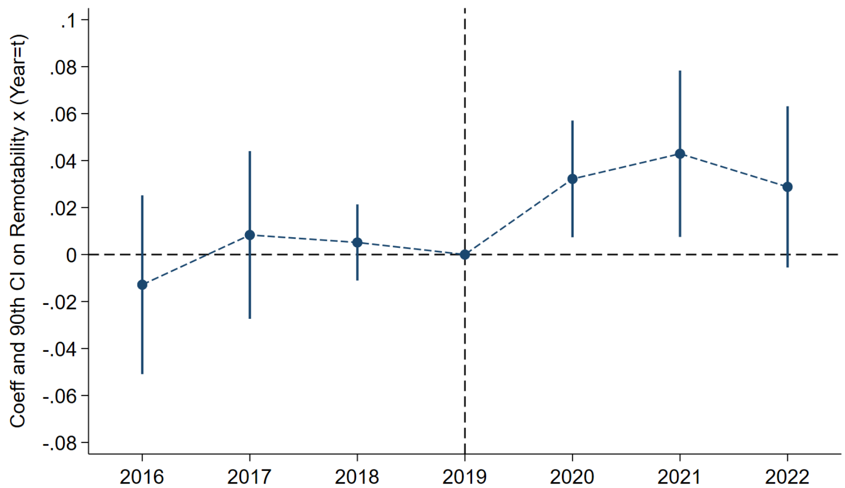
(c) Treat $= Commute$, Outcome $= SpawnShare$



(d) Treat $= Commute$, Outcome $= FounderShare$

These figures show the firm-level dynamic DID estimates. The specification follows Equations 4 and 6. The sample tracks a fixed set of employees employed in February 2020 from 2016 to 2022 regardless of whether they changed employers. We then link these individuals' spawning events (Panel A and C) and founder status (Panels B and D) to their Feb2020 employers and collapse to the Feb2020-firm-year level. We restrict to firms with employment size of 10 to 5000 as of February 2020. In Panels A and B, the treatment is ΔRW_f, the change in the Feb2020 firm's RW from 2019 to 2020/2021 average. In Panels C and D, the treatment is $Commute_f$, the average commute distance of a firm's employees in 2019. In Panels A and C (B and D), the dependent variable is 100 times the fraction of employees who spawned (who were an entrepreneur) in a given year. Each dot (bar) represents the point estimate (90th confidence interval) of the coefficient on $Treat_f \times \mathbb{1}(Year = t)$. 2019 is the omitted base year. Standard errors are clustered by the NAICS 4-digit of the Feb2020 firm.

Figure 5: Aggregate-Level Evidence: BDS New Employer Firm Entry



This figure shows how industry-level new employer firm entry changes around COVID for industries with different levels of remotability—the extent to which jobs in an industry can be performed at home or remotely (Dingel and Neiman, 2020). We estimate a dynamic DID of the following specification at the industry(NAICS 3digit)-year level:

$$Ln(no.\ of\ new\ employer\ firms)_{n,t} = \alpha_n + \beta_t + \sum_{t \neq 2019}^{2016 \to 2022} \theta_t \times Remotability_n \times \mathbb{1}(Year = t) + \epsilon_{n,t}$$

α_n indicates industry fixed effects. β_t indicates year fixed effects. The dependent variable is the log number of new employer businesses started in a NAICS-3digit-year based on US Business Dynamic Statistics (BDS). *Remotability* comes from Dingel and Neiman (2020) and is the average remotability of an industry's jobs; it is standardized before interacting with year indicators. The sample is from 2016 to 2022, the last year of BDS. The figure plots the coefficients and 90th confidence intervals of the interaction terms θ_t.

Table 1: Summary Statistics: Cross Sectional Sample

Panel A. Cross-Sectional Sample: Individual Level

Variable	N	p5	p50	p95	Mean	SD
$\text{Spawn}_{\text{post}}$	13,542,997	0.000	0.000	0.000	0.337	5.799
RW_{post}	13,542,997	0.459	0.726	0.835	0.695	0.116
Ln(emp)	13,542,997	2.996	6.084	8.295	5.903	1.673
Firm age	13,542,997	10.000	47.000	159.000	62.272	49.501
$\text{SpawnShare}_{\text{pre}}$	13,542,997	0.000	0.000	0.265	0.064	0.357
RW_{pre}	13,542,997	0.253	0.586	0.778	0.555	0.154
Tenure	13,542,997	1.000	3.000	19.000	5.206	6.250
Seniority	13,542,997	1.000	2.000	5.000	2.461	1.536
Ln(salary)	13,542,997	10.147	11.176	12.009	11.127	0.618
Prior founder	13,542,997	0.000	0.000	0.000	0.002	0.048
Commute	13,542,997	-0.621	0.289	1.361	0.317	0.621
BizClose	12,855,359	0.070	0.200	0.307	0.180	0.076

Panel B. Cross-Sectional Sample: Firm Level

Variable	N	p5	p50	p95	Mean	SD
$\text{SpawnShare}_{\text{post}}$	136,121	0.000	0.000	2.941	0.425	1.475
RW_{post}	136,121	0.417	0.713	0.865	0.685	0.136
Ln(emp)	136,121	2.303	3.296	5.894	3.599	1.125
Firm age	136,121	9.000	35.000	128.000	47.283	39.088
$\text{SpawnShare}_{\text{pre}}$	136,121	0.000	0.000	0.000	0.081	0.727
RW_{pre}	136,121	0.227	0.582	0.816	0.557	0.173
Avg. Tenure	136,121	2.312	5.212	10.556	5.656	2.597
Avg. Seniority	136,121	1.619	2.520	3.701	2.575	0.632
Avg. Ln(salary)	136,121	10.737	11.179	11.624	11.179	0.273
Avg. Prior founder	136,121	0.000	0.000	0.019	0.003	0.013
Commute	136,121	-1.152	0.160	1.282	0.121	0.754
BizClose	129,830	0.069	0.187	0.307	0.177	0.077

This table presents the summary statistics for our cross-sectional samples. The individual-level sample focuses on all Feb2020 employees of firm of employment size 10 to 5000 as of February 2020. The firm-level sample includes all firms with an employment size of 10 and 5000 as of February 2020.

Table 2: Summary Statistics: Firm Panel

Variable	N	p5	p50	p95	Mean	SD
SpawnShare	965,034	0.000	0.000	0.090	0.099	0.703
FounderShare	965,034	0.000	0.000	2.041	0.321	1.338
ΔRW	965,034	-0.118	0.124	0.391	0.128	0.154
Commute	955,297	-1.152	0.160	1.282	0.121	0.754
BizClose	920,185	0.069	0.187	0.307	0.178	0.077
Post2020	965,034	0.000	0.000	1.000	0.429	0.495

This table presents the summary statistics for our firm-level panels. The sample is based on a fixed set of Feb2020 employees employed in firms with 10 to 5000 employees in February 2020. We track these individuals over time across all their employers from 2016 to 2022 and collapse to the Feb2020-firm-year level.

Table 3: Cross-Sectional Analysis: OLS

	(1)	(2)	(3)	(4)
Sample:	Individual-level		Firm-level	
Dep var:	$Spawn_{post}$		$SpawnShare_{post}$	
RW_{post}	0.208***	0.145***	0.116***	0.113***
	(0.028)	(0.025)	(0.038)	(0.038)
Ln(emp)	-0.030***	-0.027***	-0.034***	-0.034***
	(0.003)	(0.002)	(0.004)	(0.004)
Firm age	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
$SpawnShare_{pre}$	0.130***	0.114***	0.066***	0.066***
	(0.010)	(0.009)	(0.011)	(0.011)
RW_{pre}	0.016	0.038*	0.021	0.021
	(0.023)	(0.020)	(0.038)	(0.038)
Tenure	-0.014***	-0.008***	-0.041***	-0.040***
	(0.001)	(0.000)	(0.003)	(0.003)
Seniority	0.094***	0.077***	0.197***	0.200***
	(0.007)	(0.005)	(0.018)	(0.019)
Ln(salary)	0.026***	-0.014**	0.137***	0.135***
	(0.006)	(0.006)	(0.034)	(0.033)
Prior founder	4.538***	4.386***	8.293***	8.283***
	(0.155)	(0.147)	(0.797)	(0.797)
Age		-0.007***		-0.002***
		(0.000)		(0.000)
Has grad degree		0.078***		-0.004
		(0.009)		(0.011)
Top100 BA		0.079***		0.051**
		(0.015)		(0.023)
Educ missing		-0.003		-0.018**
		(0.005)		(0.008)
NAICS 4-dig FE	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes
Job Role FE	No	Yes	No	No
Observations	13,542,997	13,526,705	136,121	136,023
R-squared	0.003	0.005	0.052	0.053

This table reports the OLS estimates of the effect of pandemic-era remote work on employees' entrepreneurial spawning. The key independent variable RW_{post} is the Feb2020 firm's average RW in 2020 and 2021. In Columns 1 and 2, the sample is at the individual level and includes all Feb2020 employees employed with firms of employment size 10 to 5000 as of February 2020. The dependent variable is 100 times a dummy indicating that the employee started a new business between March 2020 and December 2022. In Columns 3 and 4, the sample is at the firm level and includes all firm with employment size of 10 to 5000 as of February 2020. The dependent variable is 100 times the fraction of employee starting a new business between March 2020 and December 2022. Columns 1 and 3 estimate the baseline specification, while Columns 2 and 4 add additional controls. Standard errors are reported in parentheses and are clustered at the NAICS 4-digit level. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 4: Cross-Sectional Analysis: 2SLS

	(1)	(2)	(3)	(4)
Sample:	Individual-level		Firm-level	
Dep var:	$Spawn_{post}$		$SpawnShare_{post}$	
RW_{post}	3.218***	2.425***	1.308*	1.321*
	(0.878)	(0.743)	(0.677)	(0.675)
Ln(emp)	-0.041***	-0.035***	-0.039***	-0.039***
	(0.004)	(0.003)	(0.005)	(0.005)
Firm age	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
$\text{SpawnShare}_{\text{pre}}$	0.118***	0.106***	0.065***	0.065***
	(0.010)	(0.010)	(0.011)	(0.011)
RW_{pre}	-1.293***	-0.950***	-0.451	-0.457
	(0.365)	(0.311)	(0.278)	(0.277)
Tenure	-0.014***	-0.008***	-0.040***	-0.038***
	(0.001)	(0.000)	(0.003)	(0.003)
Seniority	0.091***	0.075***	0.181***	0.184***
	(0.006)	(0.005)	(0.022)	(0.022)
Ln(salary)	0.031***	-0.011*	0.172***	0.171***
	(0.007)	(0.006)	(0.041)	(0.041)
Prior founder	4.523***	4.379***	8.161***	8.150***
	(0.154)	(0.146)	(0.779)	(0.779)
Age		-0.006***		-0.001***
		(0.000)		(0.000)
Has grad degree		0.077***		-0.005
		(0.009)		(0.011)
Top100 BA		0.074***		0.044*
		(0.015)		(0.023)
Educ missing		-0.003		-0.018**
		(0.005)		(0.008)
First-stage IV coeff:				
Commute	0.009***	0.009***	0.010***	0.010***
	(0.002)	(0.002)	(0.001)	(0.001)
Kleibergen-Paap F-stat	35.0	34.9	299.5	299.9
NAICS 4-dig FE	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes
Job Role FE	No	Yes	No	No
Observations	13,542,997	13,526,705	136,121	136,023
R-squared	0.000	0.001	0.014	0.014

This table reports the 2SLS estimates of the effect of pandemic-era remote work on employees' entrepreneurial spawning. The key independent variable RW_{post} is the Feb2020 firm's average RW in 2020 and 2021. We instrument RW_{post} with $Commute$, the average commute distance of a firm's employees in 2019. In Columns 1 and 2, the sample is at the individual level and includes all Feb2020 employees employed with firms of employment size 10 to 5000 as of February 2020. The dependent variable is 100 times a dummy indicating that the employee started a new business between March 2020 and December 2022. In Columns 3 and 4, the sample is at the firm level and includes all firm with employment size of 10 to 5000 as of February 2020. The dependent variable is 100 times the fraction of employee starting a new business between March 2020 and December 2022. Columns 1 and 3 estimate the baseline specification, while Columns 2 and 4 add additional controls. Standard errors are reported in parentheses and are clustered at the NAICS 4-digit level. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 5: Firm-Level DID

Panel A. Baseline

	(1)	(2)	(3)	(4)	(5)	(6)
Dep var:		*SpawnShare*			*FounderShare*	
ΔRW × Post2020	0.008***			0.040***		
	(0.002)			(0.007)		
Commute × Post2020		0.009***			0.019***	
		(0.002)			(0.006)	
BizClose × Post2020			0.010***			0.028***
			(0.002)			(0.004)
Controls × Post2020	Yes	Yes	Yes	Yes	Yes	Yes
Feb2020-Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	965034	955297	920185	965034	955297	920185
R-squared	0.041	0.041	0.041	0.573	0.574	0.573

Panel B. Additional Industry-Year Fixed Effects

	(1)	(2)	(3)	(4)	(5)	(6)
Dep var:		*SpawnShare*			*FounderShare*	
ΔRW × Post2020	0.005**			0.024***		
	(0.002)			(0.006)		
Commute × Post2020		0.008***			0.015***	
		(0.002)			(0.005)	
BizClose × Post2020			0.008***			0.019***
			(0.001)			(0.004)
Controls × Post2020	Yes	Yes	Yes	Yes	Yes	Yes
Feb2020-Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
NAICS4dig-Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	965020	955283	920171	965020	955283	920171
R-squared	0.041	0.041	0.041	0.575	0.575	0.575

This table reports the DID results estimated on a firm-year panel following Equation 3 (Columns 1-3) and Equation 5 (Columns 4-6). Panel A estimates the baseline specification and Panel B additionally includes industry-year fixed effects. The sample tracks a fixed set of employees employed in February 2020 over 2016-2022 regardless of who their employer was in a given year. We then link these individuals' spawning events (Columns 1 to 3) and founder status (Columns 4 and 6) to their Feb2020 employers and collapse to Feb2020-firm-year level. We restrict to firms with employment size of 10 to 5000 as of February 2020. *SpawnShare* is 100 times the fraction of employees who spawned in a year. *FounderShare* is 100 times the fraction of employees who were a founder (i.e., remained in entrepreneurship) in a year. *Post2020* is a dummy indicating years 2020-2022. We interact *Post2020* with three continuous treatment variables, ΔRW, *Commute*, and *BizClose*, all standardized to reflect the effect of a one standard deviation change. All columns include firm and year fixed effects and the interactions between *Post2020* and firm-level controls. Standard errors are reported in parentheses and are clustered by the NAICS 4-digit of the Feb2020 firm. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 6: Cross-Sectional Analysis: Conditional On Turnover

	(1)	(2)
Model:	2SLS	
Sample:	Individual-level	
Dep var:	$Spawn_{post}$	
$\mathrm{RW}_{\mathrm{post}}$	7.135**	5.621*
	(3.487)	(3.256)
Ln(emp)	-0.210***	-0.184***
	(0.020)	(0.017)
Firm age	-0.002***	-0.002***
	(0.001)	(0.000)
$\mathrm{SpawnShare}_{\mathrm{pre}}$	0.465***	0.411***
	(0.046)	(0.044)
$\mathrm{RW}_{\mathrm{pre}}$	-2.417*	-1.886
	(1.449)	(1.351)
Tenure	0.045***	0.071***
	(0.008)	(0.008)
Seniority	0.501***	0.444***
	(0.031)	(0.027)
Ln(salary)	1.021***	0.890***
	(0.045)	(0.042)
Prior founder	18.844***	18.095***
	(0.604)	(0.571)
Age		-0.034***
		(0.003)
Has grad degree		0.474***
		(0.043)
Top100 BA		0.187**
		(0.077)
Educ missing		0.107***
		(0.030)
First-stage IV coeff:		
Commute	0.009***	0.009***
	(0.002)	(0.002)
Kleibergen-Paap F-stat	37.9	37.4
NAICS 4-dig FE	Yes	Yes
County FE	Yes	Yes
Job Role FE	No	Yes
Observations	2,163,871	2,160,920
R-squared	0.012	0.009

This table examines the impact of pandemic-era remote work on employees' entrepreneurial spawning at the individual level, conditional on turnover. The sample focuses on all Feb2020 employees employed with a firm of size 10 and 5000 as of February 2020, and who have left the firm between March 2020 and December 2022. The dependent variable is 100 times a dummy indicating that the employee started a new business between March 2020 and December 2022. The key independent variable RW_{post} is the Feb2020 firm's average RW in 2020 and 2021. Both columns report the 2SLS results, where RW_{post} is instrumented with $Commute$, the average commute distance of a firm's employees in 2019. Standard errors are reported in parentheses and are clustered at the NAICS 4-digit level. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 7: Quality of Marginally Spawned Firms

Panel A. Quality Conditional on Spawning

	(1)	(2)
Model:	2SLS	
Dep var:	Initial employment	VC-backed
RW_{post}	2.901*	0.918***
	(1.488)	(0.273)
% effect	38.1%	171.5%
Kleibergen-Paap F-stat	44.048	44.048
Controls	Yes	Yes
NAICS 4-dig FE	Yes	Yes
County FE	Yes	Yes
Observations	45324	45324
R-squared	-0.012	-0.087

Panel B. Spawning Response by Quality

	(1)	(2)	(3)	(4)
Model:	2SLS			
Dep var:	$Spawn_{post}$ as:			
	Employer	Non-employer	VC-backed	Non-VC-backed
RW_{post}	0.889***	2.330***	0.784***	2.435***
	(0.276)	(0.666)	(0.235)	(0.695)
% effect	184%	96%	433%	89%
P-val of diff in % effect	0.044		0.003	
Kleibergen-Paap F-stat	35.043	35.043	35.043	35.043
Controls	Yes	Yes	Yes	Yes
NAICS 4-dig FE	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes
Observations	13542997	13542997	13542997	13542997
R-squared	0.000	0.001	-0.002	0.001

This table examines the quality of the spawned firms. Panel A conditions on spawned firms and examines changes in their average quality. Panel B splits the spawning outcome by the quality of the spawned firm, hence evaluating the quality of the marginally spawned firms. The specifications are 2SLS using the instrument *Commute*. In Panel A, the sample consists of all Feb2020 employees that spawned between March 2020 and December 2022. The dependent variables are maximum employment of the spawned firm within two years of entry, and a dummy indicating whether the firm received subsequent VC backing. In Panel B, the sample consists of all Feb2020 employees. The dependent variables decompose the spawning outcome by whether the new firm is an employer or non-employer (Columns 1 and 2) and whether the new firm received subsequent VC funding (Columns 3 and 4). The key independent variable RW_{post} is the Feb2020 firm's average RW in 2020 and 2021. % effect indicates the percentage effect of a one-std-dev increase in RW_{post} on outcome relative to outcome mean. P-values indicate the significance of the % effect difference between two adjacent columns. Standard errors are in parentheses and are clustered at the NAICS 4-digit level. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 8: Heterogeneity by Experimentation Value

Panel A. Industry Risk Conditional on Spawning

	(1)
Model:	2SLS
Dep var:	Industry risk
RW_{post}	0.069**
	(0.030)
% effect	5.2%
Kleibergen-Paap F-stat	41.5
Controls	Yes
NAICS 4-dig FE	Yes
County FE	Yes
Observations	42686
R-squared	-0.047

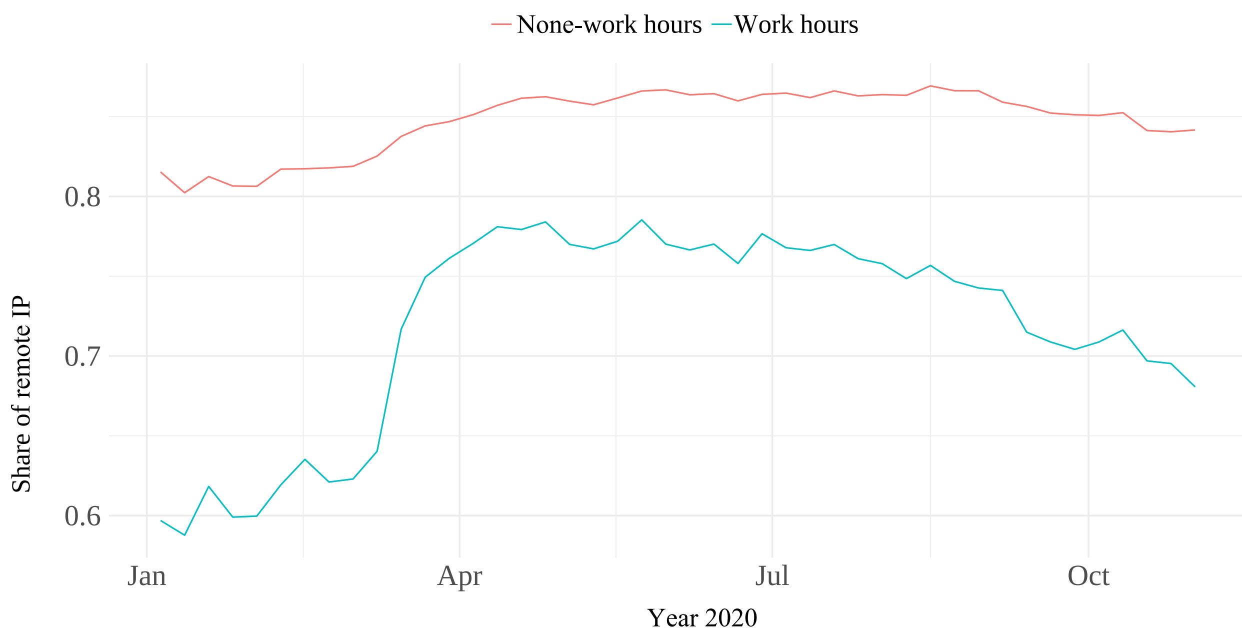
Panel B. Spawning Response by Industry Risk

	(1)	(2)
Model:	2SLS	
Dep var:	Spawn into industry with	
	High risk	Low risk
RW_{post}	1.966***	0.882**
	(0.484)	(0.425)
% effect	143.2%	64.3%
P-val of diff in % effect	0.013	
Kleibergen-Paap F-stat	35.043	35.043
Controls	Yes	Yes
NAICS 4-dig FE	Yes	Yes
County FE	Yes	Yes
Observations	13542997	13542997
R-squared	0.000	0.001

This table shows the heterogeneity of our individual-level results with respect to experimentation value of the entered industry, measured by industry young firm exit risk. The specifications follow Table 7. Panel A conditions on spawned firms and examines changes in the average exit risk of the entered industries. Panel B splits spawning response by the risk of the entered industry, hence evaluating the industry risk of the marginally spawned firms. The specifications are 2SLS using the instrument *Commute*. In Panel A, the sample consists of all Feb2020 employees that spawned between March 2020 and December 2022. The dependent variable is average exit rates of young (age$\leq$5) firms in 2015-2019 in the entered NAICS4-digit industry, based on U.S. Business Dynamic Statistics (BDS). In Panel B, the sample consists of all Feb2020 employees. The dependent variables decompose the spawning outcome by whether the spawned firm is in a NAICS4-digit industry with above or below median exit risk. The key independent variable RW_{post} is the Feb2020 firm's average RW in 2020 and 2021. % effect indicates the percentage effect of a one-standard-deviation increase in RW_{post} on outcome relative to outcome mean. P-value indicates the significance of the % effect difference between two adjacent columns. Standard errors are in parentheses and are clustered at the NAICS 4-digit level. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

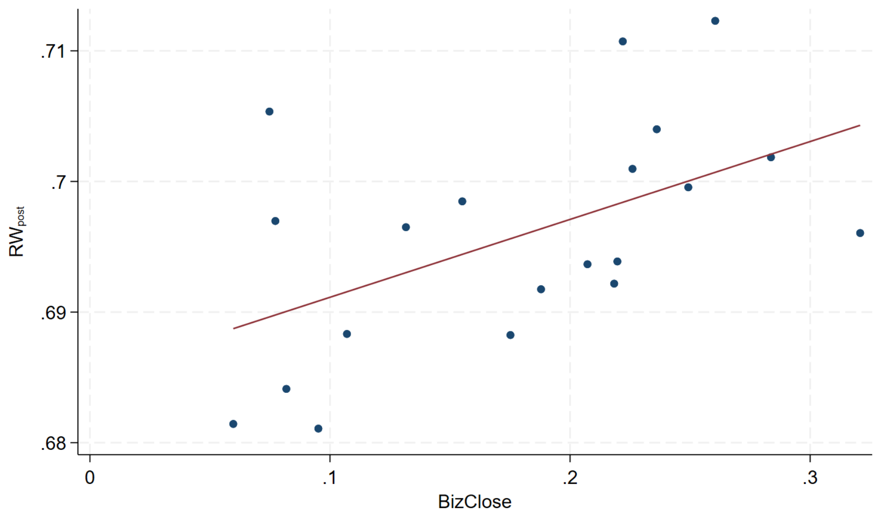
Appendix Figures and Tables

Figure A.1: IP Activity During Work Hours versus Non-Work Hours



This graph plots the weekly share of remote IP activities during work hours (Monday-Friday 9 a.m. to 6 p.m. local time) and non-work hours in year 2020. Our classification for IP addresses is described in Appendix Section A.1, involving a mixture of hand-classification and a machine learning algorithm.

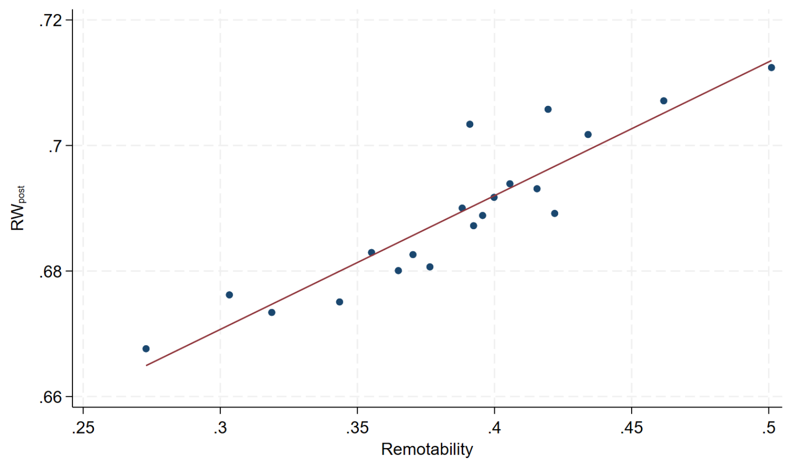
Figure A.2: RW_{post} and County-Level Business Closure Orders



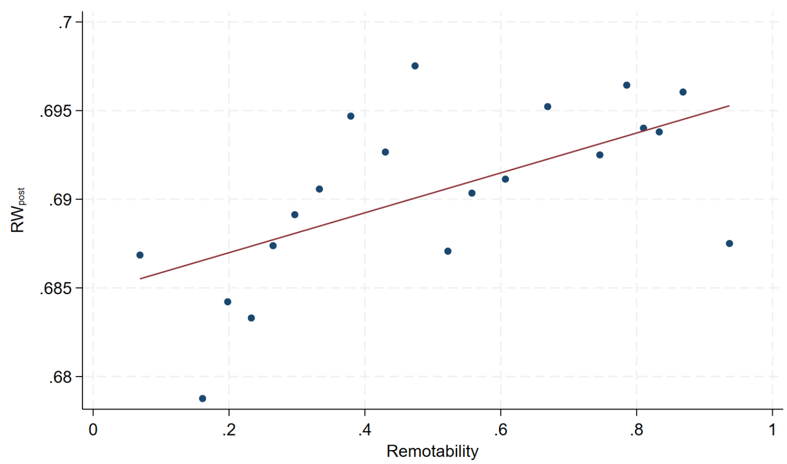
This figure shows the binned scatter plot between county-level business closures imposed during the pandemic and RW_{post} for firms headquartered in that county. Firm-level data is binned into 20 quantiles based on the value of $BizClose$, which measures the fraction of time in 2020-2021 that a county's local businesses were mandated to close. The measure accounts for partial closures and order reversals. RW_{post} is a firm's average RW in 2020 and 2021. The binned scatter plot controls for RW_{pre}, firm's average RW in 2019. The height of each dot represents the average RW_{post} within the bin.

Figure A.3: RW_{post} and CBSA-/Industry-Level Remotability

(a) CBSA Level

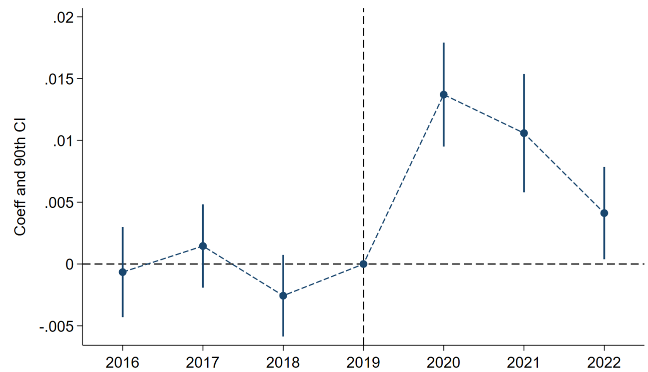


(b) NAICS-3digit Level

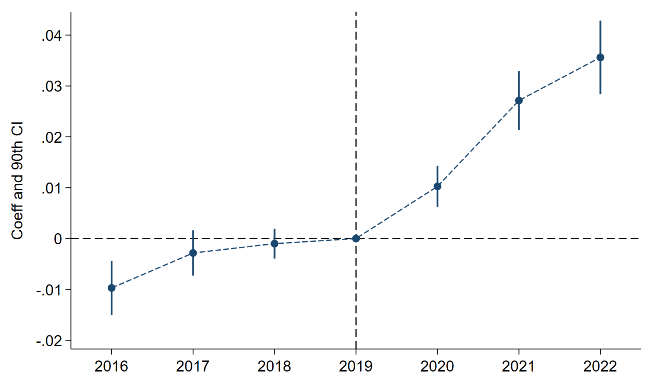


Panel A (B) shows the binned scatter plot between CBSA-level (NAICS 3digit-level) remotability and RW_{post} for firms in that CBSA (NAICS 3-digit). Firm-level data is binned into 20 quantiles based on the value of *Remotability*, which measures the fraction of jobs in a location or industry that are teleworkable (Dingel and Neiman, 2020). RW_{post} is a firm's average RW in 2020 and 2021. The binned scatter plot controls for RW_{pre}, firm's average RW in 2019. The height of each dot represents the average RW_{post} within the bin.

Figure A.4: Firm-Level DID Dynamics by $BizClose$



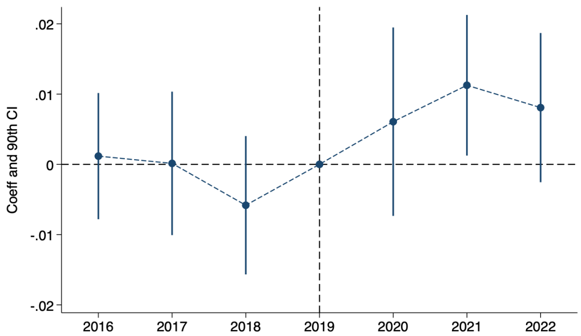
(a) Treat = $BizClose$, Outcome = $SpawnShare$



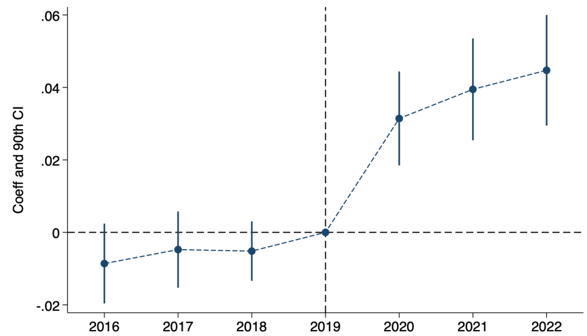
(b) Treat = $BizClose$, Outcome = $FounderShare$

This figure shows the firm-level dynamic DID effects, where the treatment is $BizClose$, the fraction of time in 2020 and 2021 that a county's local businesses were mandated to close. Panel A (B) estimates Equation 4 (Equation 6). The sample tracks a fixed set of employees employed in February 2020 from 2016 to 2022 regardless of whether they have changed employers. We then link these individuals' spawning events (Panel A) and founder status (Panel B) to their Feb2020 employers and collapse to the Feb2020-firm-year level. We restrict to firms with employment size of 10 to 5000 as of February 2020. The dependent variable is 100 times the fraction of employees who spawned (who were a founder) in a given year in Panel A (B). Each dot (bar) represents the point estimate (90th confidence interval) of the coefficient on $BizClose \times \mathbb{1}(Year = t)$. 2019 is the omitted base year. Standard errors are clustered by the NAICS 4-digit of the Feb2020 firm.

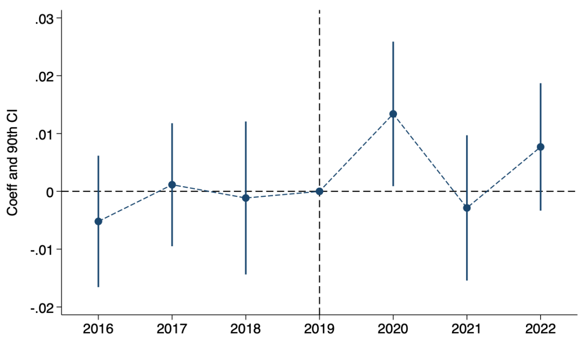
Figure A.5: Firm-Level DID Dynamics: Fixing Employees in 2016



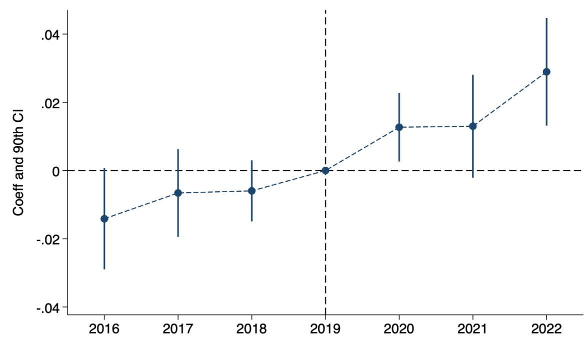
(a) Treat = ΔRW, Outcome = $SpawnShare$



(b) Treat = ΔRW, Outcome = $FounderShare$



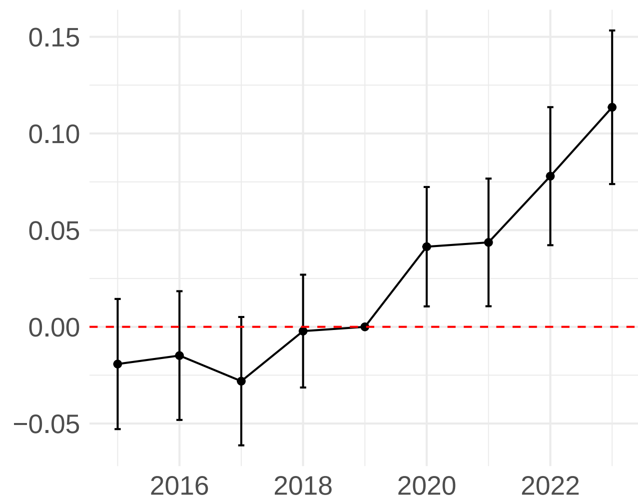
(c) Treat = $Commute$, Outcome = $SpawnShare$



(d) Treat = $Commute$, Outcome = $FounderShare$

This figure shows the firm-level dynamic DID estimates, based on a fixed set of employees employed in January 2016 and whose Jan2016 employer survived till February 2020. The specification follows Equations 4 and 6. The sample tracks a fixed set of employees employed in January 2016 from 2016 to 2022 regardless of whether they changed employers. We then link these individuals' spawning events (Panel A and C) and founder status (Panels B and D) to their Jan2016 employers and collapse to the Jan2016-firm-year level. We restrict to Jan2016 firms with employment size of 10 to 5000 in February 2020. In Panels A and B, the treatment is ΔRW_f, the change in the Jan2016 firm's RW from 2019 to 2020/2021. In Panels C and D, the treatment is $Commute_f$, the average commute distance of a firm's employees in 2019. In Panels A and C (B and D), the dependent variable is 100 times the fraction of employees who spawned (who were a founder) in a given year. Each dot (bar) represents the point estimate (90th confidence interval) of the coefficient on $Treat_f \times \mathbb{1}(Year = t)$. 2019 is the omitted base year. Standard errors are clustered by the NAICS 4-digit of the Jan2016 firm.

Figure A.6: Aggregate Evidence: Employment at New Firms

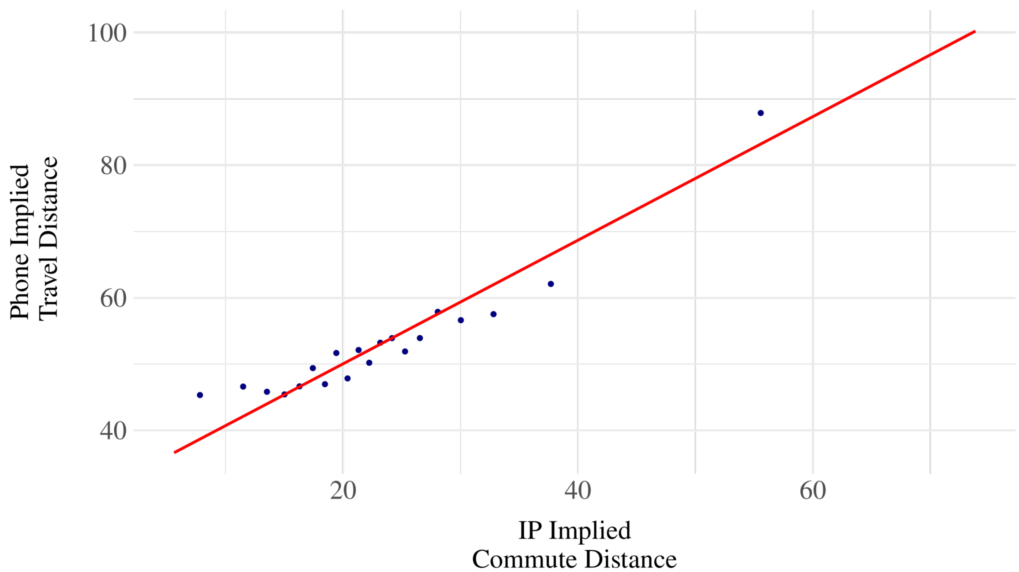


This figure shows how industry-county-level total new firm employment changes around COVID depending on an industry's remotability—the extent to which its jobs can be performed at home or remotely (Dingel and Neiman, 2020). We estimates a dynamic DID of the following specification at the NAICS2digit-county-year level, using QWI data:

$$\textit{Ln(employment at new firms)}_{n,c,t} = \alpha_{n,c} + \beta_{c,t} + \sum_{n \neq 2019}^{2015 \to 2023} \theta_t \times \textit{Remotability}_n \times \mathbb{1}(Year = t) + \epsilon_{n,c,t}$$

, where $\alpha_{n,c}$ indicates industry-county fixed effects, $\beta_{c,t}$ indicates county-year fixed effects, and $\textit{Remotability}_n$ is industry-level remotability. The dependent variable is total employment by new firms started in a NAICS-2digit-county-year based on US Quarterly Workforce Indicators (QWI). *Remotability* comes from Dingel and Neiman (2020) and is the average remotability of an industry's jobs; it is standardized before interacting with year indicators. The sample is from 2015 to 2023. 2019 is the omitted base year. The figure plots the coefficients and 90th confidence intervals of the interaction terms θ_t.

Figure A.7: Comparing Our Measure of Commute Distance to SafeGraph Mobility Data



This figure shows the binned scatter relationship between our IP-implied commute distance measure (*Commute*) and commute distance inferred from SafeGraph mobility data (both in kilometers). Appendix A.2 provides details on how we compute each measure. Firm-level data is binned into 20 quantiles based on the value of *Commute* for a firm in 2019. The height of each dot represents the average SafeGraph-based commute distance within the bin.

Table A.1: Top and Bottom Industries by ΔRW

NAICS 2-digit	Description	Average ΔRW
Top 5 industries		
51	Information	0.149
81	Other Services (except Public Administration)	0.143
54	Professional, Scientific, and Technical Services	0.141
21	Mining, Quarrying, and Oil and Gas Extraction	0.135
56	Administrative & Support, Waste Management and Remediation Services	0.134
Bottom 5 industries		
55	Management of Companies and Enterprises	0.113
62	Health Care and Social Assistance	0.112
11	Agriculture, Forestry, Fishing and Hunting	0.110
23	Construction	0.109
44	Retail trade	0.106

This table shows the top and bottom 5 NAICS 2-digit industries by the average increase in RW from 2019 to 2020/21.

Table A.2: County-Level RW Measure and Mobile Geo-Location

	(1)	(2)	(3)	(4)
		RW		
SafeGraph Worklike Behavior	-0.756^{***}		-0.743^{***}	-0.214^{***}
	(0.037)		(0.040)	(0.023)
SafeGraph Social Distancing		0.207^{***}	0.028	-0.032^{**}
		(0.022)	(0.023)	(0.014)
Measured during work hours	Yes	Yes	Yes	No
County FE	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes
Observations	272,378	272,378	272,378	275,184
R^2	0.700	0.705	0.705	0.675

This table correlates our internet activity-based remote work measure with two cell phone location-based measures from SafeGraph at the county-week level: 1) Worklike Behavior, defined as a cell phone visiting the person's "common day-time location," a place where a device is typically observed during 8am to 5pm on a workday; 2) Social Distancing, the fraction of phones that remain completely at home to the total number of devices. The specification is

$$RW_{c,w} = \alpha_c + \alpha_w + \beta SafeGraphMobility_{c,w} + \epsilon_{c,w}$$

where $RW_{c,w}$ is the fraction of remote internet activity is a county-week and $SafeGraphMobility_{c,w}$ is one of the two SafeGraph measures. Standard errors are reported in parentheses and are clustered at the county level. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table A.3: Cross-Sectional Analysis: First Stage Results

	(1)	(2)	(3)	(4)
Dep var:		RW_{post}		
Sample:	Individual-level		Firm-level	
Commute	0.009***		0.010***	
	(0.002)		(0.001)	
BizClose		0.036***		0.061***
		(0.011)		(0.006)
Ln(emp)	0.003***	0.004***	0.004***	0.005***
	(0.000)	(0.001)	(0.000)	(0.000)
Firm age	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
$\text{SpawnShare}_{\text{pre}}$	0.004***	0.005***	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
RW_{pre}	0.432***	0.435***	0.391***	0.399***
	(0.014)	(0.015)	(0.008)	(0.009)
Tenure	-0.000***	-0.000***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Seniority	0.001***	0.001***	0.012***	0.012***
	(0.000)	(0.000)	(0.001)	(0.001)
Ln(salary)	-0.002***	-0.001	-0.029***	-0.018***
	(0.001)	(0.001)	(0.003)	(0.003)
Prior founder	0.005***	0.006***	0.108***	0.126***
	(0.001)	(0.001)	(0.027)	(0.028)
Kleibergen-Paap F-stat	35.0	10.1	299.5	87.1
NAICS 4-dig FE	Yes	Yes	Yes	Yes
County FE	Yes	No	Yes	No
Observations	13,542,997	12,855,359	136,121	130,142
R-squared	0.446	0.407	0.332	0.310

The table shows the full first-stage results for our 2SLS specifications in Tables 4 and A.4, with firm-level *Commute* (Columns 1 and 3) and county-level *BizClose* (Columns 2 and 4) as instruments. Standard errors are reported in parentheses and are clustered at the NAICS 4-digit level. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table A.4: Cross-Sectional Analysis: Alternative Instrument

	(1)	(2)	(3)	(4)
Model:	OLS	2SLS	OLS	2SLS
Sample:	Individual-level		Firm-level	
Dep var:	$Spawn_{post}$		$SpawnShare_{post}$	
RW_{post}	0.262***	7.818***	0.155***	4.115***
	(0.034)	(2.496)	(0.039)	(1.018)
Ln(emp)	-0.025***	-0.057***	-0.030***	-0.052***
	(0.003)	(0.011)	(0.005)	(0.008)
Firm age	-0.001***	0.000	-0.001***	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
$SpawnShare_{pre}$	0.139***	0.103***	0.065***	0.061***
	(0.012)	(0.015)	(0.012)	(0.012)
RW_{pre}	-0.010	-3.297***	0.012	-1.567***
	(0.026)	(1.104)	(0.036)	(0.394)
Tenure	-0.015***	-0.013***	-0.044***	-0.038***
	(0.001)	(0.001)	(0.003)	(0.003)
Seniority	0.094***	0.086***	0.189***	0.143***
	(0.007)	(0.006)	(0.018)	(0.019)
Ln(salary)	0.037***	0.040***	0.185***	0.241***
	(0.006)	(0.009)	(0.032)	(0.038)
Prior founder	4.522***	4.472***	8.530***	8.016***
	(0.158)	(0.151)	(0.819)	(0.805)
First-stage IV coeff:				
BizClose		0.034***		0.060***
		(0.011)		(0.006)
Kleibergen-Paap F-stat		10.1		87.1
NAICS 4-dig FE	Yes	Yes	Yes	Yes
Observations	12,855,359	12,855,359	130,142	130,142
R-squared	0.003	-0.011	0.040	-0.069

This table shows the effect of pandemic-era remote work on employees' entrepreneurial spawning, using our alternative instrument $BizClose$. The key independent variable RW_{post} is the Feb2020 firm's average RW in 2020 and 2021. We instrument RW_{post} with $BizClose$, the fraction of time in 2020 and 2021 that a county's local businesses were mandated to close (accounting for partial closures and reversals as well). In Columns 1 and 2, the sample is at the individual level and includes all Feb2020 employees of firms of employment size 10 to 5000 as of February 2020. The dependent variable is 100 times a dummy indicating that the employee started a new business between March 2020 and December 2022. In Columns 3 and 4, the sample is at the firm level and includes on all firm with employment size of 10 to 5000 as of February 2020. The dependent variable is 100 times the fraction of employee starting a new business between March 2020 and December 2022. Columns 1 and 3 estimate the OLS results while Columns 2 and 4 estimate the 2SLS result. Standard errors are reported in parentheses and are clustered at the NAICS 4-digit level. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table A.5: Testing the Monotonicity Condition

Subsample by:	RW_{pre}	Ln(emp)	Firm age	Prior spawning ratio	Avg. tenure	Avg. seniority	Avg. ln(salary)	Avg. prior founder	Avg. grad degree	Avg. top100 BA	Avg. age
	First-stage coeff on *Commute*										
High	0.004***	0.013***	0.012***	0.011***	0.010***	0.010***	0.010***	0.013***	0.010***	0.013***	0.010***
	(0.001)	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
Low	0.014***	0.009***	0.009***	0.010***	0.010***	0.011***	0.011***	0.010***	0.011***	0.011***	0.011***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
	First-stage coeff on *BizClose*										
High	0.045***	0.051***	0.054***	0.053**	0.050***	0.071***	0.069***	0.049***	0.057***	0.065**	0.042***
	(0.006)	(0.007)	(0.008)	(0.022)	(0.009)	(0.008)	(0.009)	(0.009)	(0.014)	(0.025)	(0.009)
Low	0.076***	0.067***	0.064***	0.060***	0.065***	0.047***	0.047***	0.061***	0.061***	0.058***	0.074***
	(0.009)	(0.008)	(0.008)	(0.007)	(0.007)	(0.008)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)

The table tests the monotonicity condition for our instruments following the method in Dobbie et al. (2018). Monotonicity implies that the first-stage coefficient on the instrument should be non-negative in all subsamples formed based on observables. We test this implication in subsamples split by various firm characteristics (indicated in top row) at the median. Given our instruments are at the firm-/county-level, we test this condition in our firm-level sample. The table reports the coefficient and standard errors on our instruments in the first stage, controlling for the same variables and fixed effects as the firm-level specification in Column 3 of Table 3. Standard errors are reported in parentheses and are clustered at the NAICS 4-digit level. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table A.6: Correlation of IVs with Pre-Pandemic Spawning and Employee Founder Experience

	(1)	(2)	(3)	(4)
	$SpawnShare_{pre}$		$PriorFounderShare$	
Commute	0.001		0.007	
	(0.003)		(0.005)	
BizClose		0.029		0.030
		(0.027)		(0.055)
Ln(emp)	-0.007***	-0.007***	-0.037***	-0.037***
	(0.002)	(0.002)	(0.004)	(0.004)
Firm age	-0.000***	-0.000***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
RW_{pre}	0.013	0.014	0.131***	0.134***
	(0.010)	(0.010)	(0.028)	(0.028)
Tenure	-0.009***	-0.009***	-0.005**	-0.008***
	(0.001)	(0.001)	(0.002)	(0.002)
Seniority	0.036***	0.037***	0.132***	0.129***
	(0.007)	(0.007)	(0.017)	(0.017)
Ln(salary)	0.010	0.013	0.131***	0.155***
	(0.013)	(0.013)	(0.031)	(0.033)
NAICS 4-dig FE	Yes	Yes	Yes	Yes
County FE	Yes	No	Yes	No
Observations	136,121	131,344	136,139	131,362
R-squared	0.018	0.008	0.034	0.022

The table provides evidence consistent with our instruments satisfying the exclusion condition. We do this by showing an insignificant relationship between each of our two instruments and firms' pre-pandemic spawning share and the share of employees with past founder experience, conditional on other controls. The dependent variable in Columns 1 and 2 is 100 times the fraction of employees that spawned from the firm to become an entrepreneur in 2019. The dependent variable in Columns 3 and 4 is 100 times the fraction of employees in 2019 with past founder experience. All other control variables and fixed effects follow that of Column 3 of Table 3. Standard errors are reported in parentheses and are clustered at the NAICS 4-digit level. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table A.7: Cross-Sectional Analysis: Reduced Form Effects

	(1)	(2)	(3)	(4)
Sample:	Individual-level		Firm-level	
Dep var:	$Spawn_{post}$		$SpawnShare_{post}$	
Commute	0.029***		0.014**	
	(0.006)		(0.007)	
BizClose		0.281***		0.258***
		(0.054)		(0.062)
Ln(emp)	-0.032***	-0.025***	-0.034***	-0.030***
	(0.003)	(0.003)	(0.004)	(0.005)
Firm age	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
$\text{SpawnShare}_{\text{pre}}$	13.086***	14.002***	6.613***	6.320***
	(1.009)	(1.169)	(1.117)	(1.180)
RW_{pre}	0.096***	0.103***	0.060*	0.072**
	(0.025)	(0.029)	(0.034)	(0.031)
Tenure	-0.014***	-0.015***	-0.041***	-0.044***
	(0.001)	(0.001)	(0.003)	(0.003)
Seniority	0.094***	0.094***	0.197***	0.194***
	(0.007)	(0.007)	(0.018)	(0.018)
Ln(salary)	0.026***	0.034***	0.133***	0.166***
	(0.006)	(0.006)	(0.033)	(0.032)
Prior founder	4.539***	4.505***	8.303***	8.524***
	(0.156)	(0.153)	(0.797)	(0.772)
NAICS 4-dig FE	Yes	Yes	Yes	Yes
County FE	Yes	No	Yes	No
Observations	13,542,997	13,204,037	136,121	131,453
R-squared	0.003	0.003	0.052	0.040

The table presents the reduced form effects of our two instruments in our cross-sectional analysis. The sample and specification otherwise follow Columns 1 and 3 of Table 3. Standard errors are reported in parentheses and are clustered at the NAICS 4-digit level. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table A.8: Cross-Sectional Analysis: Alternative Samples

Panel A. All Firms with at Least 10 Employees

	(1)	(2)	(3)	(4)
Model:	OLS	2SLS	OLS	2SLS
Sample:	Individual-Level		Firm-Level	
Dep var:	$Spawn_{post}$		$SpawnShare_{post}$	
RW_{post}	0.210***	2.441***	0.116***	1.303*
	(0.030)	(0.873)	(0.038)	(0.674)
Kleibergen-Paap F-stat		19.112		298.816
Controls	Yes	Yes	Yes	Yes
NAICS 4-dig FE	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes
Observations	17,491,541	17,491,541	136,491	136,491
R-squared	0.003	0.540	0.052	0.014

Panel B. Restrict to Firms with Growing Employment

	(1)	(2)	(3)	(4)
Model:	OLS	2SLS	OLS	2SLS
Sample:	Individual-Level		Firm-Level	
Dep var:	$Spawn_{post}$		$SpawnShare_{post}$	
RW_{post}	0.328***	2.567***	0.175**	2.611**
	(0.053)	(0.870)	(0.081)	(1.074)
Kleibergen-Paap F-stat		25.578		79.864
Controls	Yes	Yes	Yes	Yes
NAICS 4-dig FE	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes
Observations	4,313,397	4,313,397	26984	26984
R-squared	0.004	0.002	0.096	-0.013

Table A.8: Cross-Sectional Analysis: Alternative Samples (Continued)

Panel C. Drop Top 60 NAICS-4-Digits that Boomed During COVID

	(1)	(2)	(3)	(4)
Model:	OLS	2SLS	OLS	2SLS
Sample:	Individual-Level		Firm-Level	
Dep var:	$Spawn_{post}$		$SpawnShare_{post}$	
$\mathrm{RW}_{\mathrm{post}}$	0.192***	3.559***	0.107***	1.787**
	(0.030)	(1.157)	(0.037)	(0.769)
Kleibergen-Paap F-stat		24.682		206.771
Controls	Yes	Yes	Yes	Yes
NAICS 4-dig FE	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes
Observations	11,375,140	11,375,140	113,894	113,894
R-squared	0.003	0.000	0.052	0.004

This table examines the robustness of our cross-sectional results to alternative samples. The specifications follow Columns 1 and 3 of Tables 3 and 4. The instrument is *Commute*—the average commute distance of a firm's employees in 2019. Panel A includes all firms with at least 10 employees in February 2020 (including those with more than 5000 employees). Panel B restricts to firms that experienced continued yearly employment growth during COVID from 2020 to 2022. Panel C drops the top 60 NAICS-4-digit industries (about 20% of all NAICS-4-digits in our sample) that experienced the highest aggregate employment growth during 2020-2022 relative to 2018-2019 based on Census BDS data. In all panels, Columns 1 and 2 present individual-level results based on all Feb2020 employees of these firms, and Columns 3 and 4 present collapsed firm-level results. The dependent variable is 100 times a dummy indicating (the fraction of) employee starting a new business between March 2020 and December 2022 in Columns 1 and 2 (Columns 3 and 4). The key independent variable RW_{post} is a firm's average RW in 2020 and 2021. For brevity, we do not report the coefficients on the control variables. Standard errors are reported in parentheses and are clustered at the NAICS 4-digit level. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table A.9: Firm-Level DID: First Stage Results

	(1)	(2)
Dep var:	*RW*	
Commute × Post2020	0.011***	
	(0.001)	
BizClose × Post2020		0.006***
		(0.000)
Controls × Post2020	Yes	Yes
Feb2020-Firm FE	Yes	Yes
Year FE	Yes	Yes
Observations	409413	394365
R-squared	0.665	0.664

This table shows the "first-stage" results of our firm panel analysis. The sample period is 2019 to 2021, the period for which we have firm-year level *RW* measure. The dependent variable is a firm's *RW* in a given year, measured by the fraction of employee Internet activities originating from remote IP addresses. *Post2020* is a dummy indicating years 2020 and later. We interact *Post2020* with two continuous treatment variables *Commute*, and *BizClose*, both standardized to reflect the effect of a one standard deviation change. Standard errors are reported in parentheses and are clustered at the NAICS 4-digit level. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table A.10: Cross-Sectional Analysis: Spawning Before vs After Departure

Panel A. Spawning Before Leaving Wage Job

	(1)	(2)	(3)	(4)
Model:	OLS	2SLS	OLS	2SLS
Sample:	Individual-Level		Firm-Level	
Dep var:	$Spawn\_BeforeLeaving$		$SpawnShare\_BeforeLeaving$	
RW_{post}	0.068***	0.211	0.062***	0.140
	(0.014)	(0.269)	(0.022)	(0.391)
Kleibergen-Paap F-stat		35.043		299.455
Controls	Yes	Yes	Yes	Yes
NAICS 4-dig FE	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes
Observations	13542997	13542997	136121	136121
R-squared	0.002	0.002	0.029	0.010

Panel B. Spawning After Leaving Wage Job

	(1)	(2)	(3)	(4)
Model:	OLS	2SLS	OLS	2SLS
Sample:	Individual-Level		Firm-Level	
Dep var:	$Spawn\_AfterLeaving$		$SpawnShare\_AfterLeaving$	
RW_{post}	0.140***	3.007***	0.054*	1.168**
	(0.021)	(0.782)	(0.031)	(0.497)
Kleibergen-Paap F-stat		35.043		299.455
Controls	Yes	Yes	Yes	Yes
NAICS 4-dig FE	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes
Observations	13,542,997	13,542,997	136,121	136,121
R-squared	0.002	-0.001	0.042	0.004

This table splits the dependent variable in our cross-sectional analysis by whether spawning happens before (Panel A) or after (Panel B) a worker formally leaves her Feb2020 wage employment job. The specifications follow Columns 1 and 3 of Tables 3 and 4. The instrument *Commute* is the average commute distance of a firm's employees in 2019. The key independent variable RW_{post} is a firm's average RW in 2020 and 2021. For brevity, we do not report the coefficients of the control variables. Standard errors are reported in parentheses and are clustered at the NAICS 4-digit level. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table A.11: Classified IPs Summary Statistics

Label	IPs (millions)	Events (millions)
Business	32.9	$3,580.1$
Residential	475.0	$13,189.2$
Mobile	6.7	$3,348.9$
VPN	2.8	194.9

This table presents summary statistics describing the average number of IPs and events per month by IP address type.

Table A.12: Cross-Sectional Analysis: Alternative *RW* Based on Residential IP Only

	(1)	(2)	(3)	(4)
	OLS	2SLS	OLS	2SLS
Sample:	Individual-Level		Firm-Level	
Dep var:	$Spawn_{2020\text{-}2022}$		$SpawnShare_{2020\text{-}2022}$	
$\text{RW\_residential}_{\text{post}}$	0.202***	2.281***	0.088**	1.056*
	(0.026)	(0.542)	(0.036)	(0.543)
Kleibergen-Paap F-stat		64.0		374.1
Controls	Yes	Yes	Yes	Yes
NAICS 4-dig FE	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes
Observations	13,542,997	13,542,997	136,121	136,121
R-squared	0.003	0.001	0.052	0.016

This table shows the robustness of our main result to an alternative RW measure based on remote traffic from residential IPs only (i.e., excluding mobile and VPN IPs). The specifications follow Columns 1 and 3 of Tables 3 and 4. The instrument *Commute* is the average commute distance of a firm's employees in 2019. The key independent variable $RW\_residential_{\text{post}}$ is a firm's average fraction of employee IP traffic in 2020 and 2021 that come from residential IP relative to business IP. For brevity, we do not report the coefficients on the control variables. Standard errors are reported in parentheses and are clustered at the NAICS 4-digit level. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table A.13: Cross-Sectional Analysis: Within-Firm Variation Across Establishments

	(1)	(2)
Sample:	Individual-level	
Dep var:	$Spawn_{post}$	
Worker's local BizClose	0.189***	0.116***
	(0.041)	(0.039)
Tenure	-0.011***	-0.006***
	(0.001)	0.000
Seniority	0.089***	0.074***
	(0.007)	(0.006)
Ln(salary)	0.021***	-0.019***
	(0.006)	(0.006)
Prior founder	4.106***	3.991***
	(0.202)	(0.196)
Age		-0.006***
		(0.000)
Has grad degree		0.096***
		(0.007)
Top100 BA		0.043**
		(0.018)
Educ missing		0.000
		(0.007)
Firm FE	Yes	Yes
Worker role FE	No	Yes
Observations	7,196,339	7,194,384
R-squared	0.030	0.031

This table exploits variation in workers' locations within the same employer that has multiple establishments. We test the reduced form effect of workers' local county business closure orders in 2020 and 2021 (*Worker's local BizClose*) on their subsequent spawnings, controlling for firm fixed effects. The sample is at the individual level, for individuals with non-missing Feb2020 location reported on LinkedIn. The specifications follow Columns 1 and 2 of Table 3 but include firm fixed effects. As such, all firm-level controls are absorbed. Standard errors are reported in parentheses and are clustered by employer firm's NAICS 4-digit industry. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table A.14: Cross-Sectional Analysis: Worker Location Fixed Effects

	(1)	(2)
Model:	OLS	2SLS
Dep var:	$Spawn_{post}$	
$\mathrm{RW}_{\mathrm{post}}$	0.154***	2.091***
	(0.025)	(0.787)
Ln(emp)	-0.029***	-0.036***
	(0.003)	(0.003)
Firm age	-0.001***	-0.000***
	(0.000)	(0.000)
$\mathrm{SpawnShare}_{\mathrm{pre}}$	0.117***	0.110***
	(0.012)	(0.011)
$\mathrm{RW}_{\mathrm{pre}}$	0.043*	-0.799**
	(0.024)	(0.336)
Tenure	-0.012***	-0.012***
	(0.001)	(0.001)
Seniority	0.068***	0.067***
	(0.005)	(0.005)
Ln(salary)	-0.018***	-0.015**
	(0.006)	(0.006)
Prior founder	4.169***	4.163***
	(0.183)	(0.182)
First-stage IV coeff:		
Commute		0.009***
		(0.002)
Kleibergen-Paap F-stat		31.6
NAICS 4-dig FE	Yes	Yes
Firm County FE	Yes	Yes
Worker county FE	Yes	Yes
Observations	7,694,567	7,692,483
R-squared	0.006	0.001

This table demonstrate the robustness of our cross-sectional results to the inclusion of worker county fixed effects. The sample is at the individual level, for individuals with non-missing Feb2020 location reported on LinkedIn. The specification follows Column 1 of Tables 3 and 4, but with additional worker county fixed effects. Standard errors are reported in parentheses and are clustered by employer firm's NAICS 4-digit industry. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table A.15: Cross-Sectional Analysis Conditional On Turnover: OLS

	(1)	(2)
Model:	OLS	
Sample:	Individual-level	
Dep var:	$Spawn_{post}$	
RW_{post}	0.800***	0.496***
	(0.162)	(0.151)
Ln(emp)	-0.188***	-0.167***
	(0.018)	(0.016)
Firm age	-0.003***	-0.003***
	(0.000)	(0.000)
$\text{SpawnShare}_{\text{pre}}$	0.490***	0.429***
	(0.045)	(0.043)
RW_{pre}	0.207	0.231**
	(0.126)	(0.106)
Tenure	0.043***	0.070***
	(0.008)	(0.008)
Seniority	0.511***	0.451***
	(0.031)	(0.027)
Ln(salary)	1.013***	0.884***
	(0.043)	(0.041)
Prior founder	18.872***	18.109***
	(0.606)	(0.573)
Age		-0.034***
		(0.003)
Has grad degree		0.474***
		(0.043)
Top100 BA		0.199**
		(0.077)
Educ missing		0.110***
		(0.030)
NAICS 4-dig FE	Yes	Yes
County FE	Yes	Yes
Job Role FE	No	Yes
Observations	2,163,871	2,160,920
R-squared	0.018	0.025

This table shows the OLS estimates of the effect of pandemic-era remote work on employees' entrepreneurial spawning at the individual level, conditional on turnover. The sample and specification follow Table 6. The sample includes all Feb2020 employees of firms between size 10 and 5000 as of February 2020, and who have left the firm between March 2020 and December 2022. The dependent variable is 100 times a dummy indicating that the employee started a new business between March 2020 and December 2022. The key independent variable RW_{post} is the Feb2020 firm's average RW in 2020 and 2021. Standard errors are reported in parentheses and are clustered at the NAICS 4-digit level. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table A.16: Quality and Industry Risk of Marginally Spawned Firms: OLS

Panel A. Quality and Industry Risk Conditional on Spawning

	(1)	(2)	(3)
Model:		OLS	
Dep var:	Initial emp.	VC-backed	Industry risk
RW_{post}	0.071	0.052***	0.007***
	(0.095)	(0.014)	(0.002)
% effect	0.9%	9.7%	0.5%
Controls	Yes	Yes	Yes
NAICS 4-dig FE	Yes	Yes	Yes
County FE	Yes	Yes	Yes
Observations	45,324	45,324	42,686
R-squared	0.041	0.067	0.071

Panel B. Spawning Response by Quality and Industry Risk

	(1)	(2)	(3)	(4)	(5)	(6)
Model:	OLS	OLS	OLS	OLS	OLS	OLS
Dep var:	Spawn as		Spawn as		Spawn into industry with	
	Employer	Non-employer	VC-backed	Non-VC-backed	High risk	Low risk
RW_{post}	0.050***	0.159***	0.036***	0.173***	0.142***	0.043***
	(0.011)	(0.023)	(0.007)	(0.024)	(0.021)	(0.016)
% effect	10.3%	6.6%	19.9%	6.3%	10.3%	3.1%
P-val of diff in % effect	0.087		0.001		0.000	
Kleibergen-Paap F-stat						
Controls	Yes	Yes	Yes	Yes	Yes	Yes
NAICS 4-dig FE	Yes	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	13,542,997	13,542,997	13,542,997	13,542,997	13,542,997	13,542,997
R-squared	0.001	0.003	0.001	0.003	0.002	0.002

This table shows the OLS results for Tables 7 and 8, where we examine the quality and industry risk of spawned businesses. Panel A conditions on spawned firms and examines changes in the average quality and industry risk of the spawned firms. Panel B splits the spawning outcome by the quality/industry risk of the spawned firm, hence evaluating the quality/industry risk of the marginally spawned firms. In Panel A, the sample consists of all Feb2020 employees that spawned between March 2020 and December 2022. In Panel B, the sample consists of all Feb2020 employees. The key independent variable RW_{post} is the Feb2020 firm's average RW in 2020 and 2021. % effect indicates the percentage effect of a one-standard-deviation increase in RW_{post} on outcome relative to outcome mean. P-value indicates the significance of the % effect difference between two adjacent columns. Standard errors are in parentheses and are clustered at the NAICS 4-digit level. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table A.17: Heterogeneity by RW of Spawned Firms

	(1)	(2)	(3)	(4)
Model:	OLS	OLS	2SLS	2SLS
Dep var:	Spawned and new firm has			
	High RW	Low RW	High RW	Low RW
RW_{post}	0.017***	0.016***	0.232**	0.295***
	(0.005)	(0.005)	(0.110)	(0.111)
% effect	10%	10%	137%	177%
P-val of diff in % effect	0.859		0.668	
Kleibergen-Paap F-stat			35.043	35.043
NAICS 4-dig FE	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes
Observations	13,542,997	13,542,997	13,542,997	13,542,997
R-squared	0.000	0.000	0.000	0.000

This table shows the heterogeneity of our individual-level cross-sectional results by whether the spawned business had an above-median (high) or below-median (low) level of RW post 2020. The specification follows Column 1 of Tables 3 and 4. The instrument *Commute* is the average commute distance of a firm's employees in 2019. The dependent variable is 100 times a dummy indicating that the employee started a new business between March 2020 and December 2022 and the new business had an above- or below-median level of RW post 2020. % effect indicates the percentage effect of a one-standard-deviation increase in RW_{post} on outcome relative to outcome mean. P-value indicates the significance of the % effect difference between two adjacent columns. Standard errors are reported in parentheses and are clustered at the NAICS 4-digit level. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table A.18: Spawning in Same vs Different Industry

	(1)	(2)	(3)	(4)
	OLS	OLS	2SLS	2SLS
Dep var:	Spawn into			
	Same industry	Diff industry	Same industry	Diff industry
RW_{post}	0.032***	0.176***	0.533*	2.685***
	(0.011)	(0.027)	(0.290)	(0.797)
% effect	53%	64%	884%	969%
Kleibergen-Paap F-stat			35.043	35.043
Controls	Yes	Yes	Yes	Yes
NAICS 4-dig FE	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes
Observations	13542997	13542997	13542997	13542997
R-squared	0.002	0.003	0.000	0.000

This table decomposes our individual-level cross-sectional results by whether the spawned business is the same NAICS 3-digit industry as the original employer. The specification follows Column 1 of Tables 3 and 4. The instrument *Commute* is the average commute distance of a firm's employees in 2019. The dependent variable is 100 times a dummy indicating that the employee started a new business between March 2020 and December 2022 in the same or different NAICS 3-digit as the employer firm. % effect indicate the percentage effect of a one-standard-deviation increase in RW_{post} on outcome relative to outcome mean. Standard errors are reported in parentheses and are clustered at the NAICS 4-digit level. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table A.19: Heterogeneity by Local K12 Schools' Learning Model

	(1)	(2)
Dep var:	$Spawn_{post}$	
Sample:	Age 30-45	Other ages
RW_{post}	0.163***	0.216***
	(0.043)	(0.040)
RW_{post} × %SchoolInPerson	0.155**	-0.038
	(0.077)	(0.061)
Controls	Yes	Yes
NAICS 4-dig FE	Yes	Yes
County FE	Yes	Yes
N	6,118,981	7,162,472
R-sq	0.004	0.003

This table shows the heterogeneity of our individual-level cross-sectional results by local K12 school learning model. The interaction variable, $\%SchoolInPerson$, is the fraction of time in 2020-2021 that K12 schools in a county was in-person as opposed to remote or hybrid. The data come from `www.covidschooldatahub.com` (Jack et al., 2023). The specification follows Column 1 of Table 3. Column 1 restricts to individuals of age 30 to 45, while column 2 restricts to individuals of other ages (i.e, <30 or >45). The dependent variable is 100 times a dummy indicating that the employee started a new business between March 2020 and December 2022. Standard errors are reported in parentheses and are clustered at the NAICS 4-digit level. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table A.20: Machine Learning Classifier Summary Statistics

Panel A: Summary statistics for training sample

Label	% Top Domain	City State	State to N	% Mobile	Outdegree	Profiles per day	% Topic total	Events total
Business	0.94	2.07	0.01	0.19	$1,169.76$	5.83	13.10	$1,154.03$
Business VPN	0.68	4.61	0.01	0.16	$5,437.72$	25.63	14.70	$5,252.18$
Residential	0.22	1.51	0.04	0.72	106.91	1.75	10.81	216.43
Residential VPN	0.17	2.37	0.06	0.62	424.34	3.62	5.97	586.13
Mobile	0.15	98.97	0.19	0.93	$13,388.15$	145.86	8.48	$22,960.17$

Panel B: Machine learning accuracy statistics

	Business	Business VPN	Mobile	Residential	Residential VPN
Sensitivity	0.8991	0.7310	0.9743	0.9765	0.9024
Specificity	0.9595	0.9782	0.9856	0.9802	0.9955
Balanced Accuracy	0.9293	0.8546	0.9800	0.9783	0.9489

This table reports summary statistics for our machine learning classification of IP addresses. Panel A provides the summary statistics for our training sample. Panel B provides various accuracy statistics for our classification.

A.1 Appendix: Methodology to Classify IPs

We distinguish office IPs from remote IPs which we define as residence, VPN, or mobile IP. We make these classifications for all IP addresses observed in the data set from January 2020 to August 2020. Our classification is based on pre-pandemic information.[28] We implement two classification methods: a rule-based classification and a simple machine-learning based classification model. The rule-based approach covers approximately one fifth of the IPs but over 60% of traffic, and the machine learning-based approaches cover the remainder.

A.1 Rule-Based Approach

We obtain various data points on the entire IPv4 space at different points in time. Currently, the Internet operates on two protocols: IPv4 and IPv6. IPv4 was introduced by the Defense Advanced Research Projects Agency (DARPA) in 1981. The IPv4 protocol has 2^{32} IP addresses, or just over 4 billion. IPv6 has 2^{64} address bits and is reverse compatible with IPv4 addresses. One minor limitation worth noting is that we only have IPv4 addresses in our database, but such IP addresses are likely dominant in the dataset. According to this Network World article,[29] only 30% of IP addresses are reachable via IPv6 as of 2020.[30] At a minimum, to get online, the owner of an IP address must register information with the Regional Internet Registry (RIR). All websites have an IP address and a forward DNS record is necessary to resolve a website such as *microsoft.com* to an IP. Given that an IP address must be registered to an Internet Service Provider (ISP), we can learn the ISP, and for a large fraction of IPs, the owner of the IP address names itself. For example, the IP address 147.8.240.42 identifies itself as *vpn2fa.hku.hk*.[31]

Another option we explored is port-scan data, by which one pings IP addresses directly on various ports. For example, a computer may have an open remote desktop protocol (port 3389, typically) or mail service (sMTP on port 25). The absence of a response may mean that the computer is offline, is not listening on that port, or is not listening to the IP address making contact on that port. The responses may indicate the presence of services that are unlikely to be observed on a residential network but more likely an office or corporate network, such as a corporate mail server. We receive researcher access to data from Rapid7 and Censys, two leading cybersecurity

[28]Some IP addresses show up later on without pre-pandemic information in our dataset, but are a small fraction of total internet activity. Implicitly, the assumption we make is that the classification of an IP address to a particular type remains stable during the pandemic, or that any change in the usage of an IP address is not systematic in a way that introduces bias.

[29]https://www.networkworld.com/article/3254575/what-is-ipv6-and-why-aren-t-we-there-yet.html

[30]In addition, we surmise that these statistics about IPv6 adoption likely do not count the subset of IPv6 that are also IPv4 addresses, suggesting that uniquely IPv6 addresses that are not IPv4 are likely even smaller.

[31]The incentive of organizations to name themselves varies, but self-identifying IP addresses helps traffic from IP addresses to be recognized as legitimate. It is also desirable for a website to have a domain name instead of its website simply being a series of integers.

research firms that conduct regular censuses of the IPv4 space. From Rapid7, we retrieve the reverse Domain Name System (DNS) table every week from 2019 through 2020. We also obtain data on ports open from Censys on various common services, such as databases (MySQL, PostgreSQL), mail servers, and remote access protocols.

We also collect data on IP addresses that are used as part of residential proxy networks. Residential proxies are typically residential IP addresses where in exchange for a free service—most commonly, the users in these networks obtain a free VPN—the companies operating residential networks are allowed to pass a small amount of traffic through them. These networks are used by consumers for VPN and by corporations to study test websites from the end users' vantage point or to scrape data. We classify residential proxy IP addresses as residential, and further infer that /24 blocks (e.g., all IP addresses with the same last octet of an IP address) are also residential.[32] We obtain data from two sources. Luminati.io granted permission for us to record IP addresses on its network. Supplemental to Luminati.io, we also obtain data from Mi et al. (2019), which provides a snapshot of residential proxy networks the study procured through the middle of 2018. This provides valid data to the extent that ISPs continue to hold onto the same residential blocks two years later (even if the IPs are reassigned among household and other customers).

We classify IP addresses according to several criteria using information as of February 2020. First, we define a remote IP as a VPN if (1) its reverse DNS entry includes VPN in the subdomain but not the domain. That is, vpn.hku.hk would be a VPN, while hku.vpnservices.hk would not be. Second, we define an IP as an office location if it is consistently rated as a business per the Data Partner's algorithm. The Data Partner uses its proprietary dataset to classify IP addresses based on the temporal patterns of content consumed, the device types observed (desktop, mobile phone, tablet), and scale features (hundreds, thousands, or millions of reads from the IP). The IP classification model which it uses commercially to identify IP addresses of small businesses and large corporations is based on a ground truth set of Fortune 1000 companies' offices. Based on their historical data, we classify a business as one in which, as of February 2020, the Company has consistently identified the IP address as belonging to a business over the past 14 months. Third, we define a network as a mobile network if it satisfies two of the following criteria: (1) over the course of 2019 it had at least 50% mobile users, or if during the first two months of 2020 it had 90% mobile users, (2) MaxMind's connection type database classifies it as a mobile connection, (3) the reverse DNS of mobile network resolves to a known mobile carrier. Of course, many mobile carriers offer landlines as well. Fourth, if an IP address is not classified by any of the above criteria, we classify it as residential if it belongs in the same /24 block (e.g., the 256 addresses denoted XXX possible within a 3 octet prefix a.b.c.XXX) as a residential proxy.

[32] As argued by Mi et al. (2019), typically ISPs register /24 blocks as a bundle.

A.2 Machine Learning Approach

We classify the remainder of IP addresses using a simple machine learning model. We construct a ground truth dataset (data where we know the actual classification of each IP address) of 10,000 IP addresses.[33] To classify IP addresses, we utilize five sets of features: (1) port-scans, whether specific ports are open on the device, (2) content-based filters, which ask whether the IP address consumes more work-related content, (3) scale features, which ask what is the scale of reading activity at the IP address in terms of article reads and browsers observed (4) temporal features based on time of day of activity, and (5) geographic features, which characterize the *geographic distribution* of profiles. For example, if an IP address visitors are associated with several different states or countries, one can reasonably infer it is a mobile network or VPN and not a single household. The majority of these features are based on the reading data set, but port scans and DNS records are collected by third-party cybersecurity firms (Censys and Rapid 7). We report summary statistics of the labeled data sets in Appendix Table A.20. We sample down to reduce imbalance in the panel and execute a random forest model against the data, splitting it 70-30. We report additional details on our machine learning metrics in Section A.2.1 of the Appendix. The classification accuracy across five classes using a random forest model is around 90%, with high sensitivity and specificity for all classes.

Summary statistics of IPs. In Table A.11 we report summary statistics of classified IP addresses. Based on our classification, the majority of IP addresses are residential. Of approximately 760 million IP addresses observed, 13 billion events per month in 2020 originate from 475 million residential IP addresses. The next largest category, by events, is mobile phones, followed by businesses. VPNs are the smallest category, with a small fraction of content originating from these IPs. Although residential IPs are the largest share of IP addresses we observe, they are the least active. For each residential IP, we observe roughly 36 events per month. Other types of IP addresses, such as VPNs which are likely shared by multiple users, have double the events per IP. Meanwhile, business IPs are even larger, measured by unique users, which is consistent with business establishments being much larger on average than households.

[33] For residential labels, we use Time Warner and Comcast IP address blocks. For business addresses, the data provider has a set of known IP addresses from the Fortune 1000 which they certify are corporate IPs. For mobile networks, we define mobile networks as US mobile address networks where at least half of the observed events in the Event data set in 2019 were mobile, Maxmind's connection-type database lists it as a mobile connection, and the reverse DNS entry in February 2020 indicates it as part of the mobile network of a major U.S. telecom (e.g., AT&T). We classify VPNs of two types: residential proxies, which tend to service IP addresses from all of the world across a wide variety of users (who may originate from different companies), and business VPNs, which are generally used by members of one company. Empirically, these two types of VPNs will have different behaviors, which we separately model.

A.2.1 Details on machine learning classifier

We describe the approach we use to categorize IP addresses. We aim to model the following labels:

- Business IP: Our ground truth labels come from office IPs maintained by the business provider. However, this may capture an IP address which behaves like business IP prior to the pandemic (such as a home office used before the pandemic).
- Business VPN: Different than a consumer VPN, this VPN is used by businesses to access work-relevant content.
- Residence: A home IP address.
- Residential VPN: A home IP address that belongs to a residential proxy network. We presume a user on this type of network is not working in an office. Distinguishing between this and a normal residential IP serves only to aid modeling, so as to reduce confusion between VPNs for business, VPNs for residential use, residential IPs more broadly, and mobile IPs.
- Mobile network: A mobile network used by many different phones.

We label data from the following sources. For business IPs, we use a list of office IPs maintained by the data provider. Business VPNs are those who are classified consistently in 2019 as VPNs by the company's business IP prediction algorithm (90% of 26 weeks) and the reverse DNS entry of the VPN indicates it is a VPN in the sub-domain (e.g., vpn2fa.hku.hk). The ground truth for residential IPs is based off of Charter and Time Warner IP address blocks maintained by the company. Mobile IPs are those which across 2019 had over 200 mobile users per day. Residential VPNs were those who showed up in the dataset provided by Mi et al. (2019).

With these labels, we build features to make predictions. We do multi-class prediction. The features that are entered into the model are as follows:

- *Internet census data*: The number of ports open, the specific common ports open collected by Censys, such as MySQL. The data are taken from the first week of February.
- *Temporal patterns:* From the data provider's historical dataset, we can observe the patterns of when content is observed. The data provider calculates the fraction of content observed during work hours versus local time. We take the average of these values across 2019 and the first two months of 2020.
- *Device type features*: Browser requests are made in a format that fits a browser device) that belong to a particular OS, which implies the device type. For example, the user agent

$Mozilla/5.0(iPhone;CPUiPhoneOS12_2likeMacOSX)$ is a prefix of a user agent that implies an iPhone user.

- *Scale-based features*: The amount of content consumed at an IP address. At a network shared by many people, presumably the amount of content consumed will be larger. The device type is implied by the user agent, and the data provider tracks the fraction of user agents belonging to mobile devices versus desktops.

- *Profile-based features*: These are derived for our study. At the monthly level—in turn aggregated across our historical period of 14 months prior to March 2020—we calculate (1) the share of the top profile in an IP. In a residential IP, the share of the top profile should be larger than in an organization or mobile network. (2) the geographic dispersion of profiles (distinct cities, states, and metros associated with profiles). The novel intuition is that mobile and VPN networks are likely to come from disperse locations. (3) The fraction of profiles with an associated firm, which should characterize business networks. (4) The graph degree.

For our labeled dataset, we report summary statistics in Table A.20 below. The fraction of profiles observed on a business VPN belonging to the top domain is 94%. For a business VPN, it is 68%. This suggests either the business VPN users are either un-matched or the VPN is perhaps shared across organizations. Conversely, at residential and mobile VPNs the fraction of users associated with the domain drops dramatically. Relative to Businesses, business VPNs are generally more geographically disperse, coming from a larger percentage of cities. Similarly, relative to regular residential addresses, residential VPN users come from a wider range of cities. Mobile networks and VPNs operate at a much larger scale than our residential or business labels, given that pool more users from a broader geography. This is reflected in the final two columns. First, profiles per day reveals approximately the number of users of an IP address (assuming that user stays within a single browser) and second, the total number of events reveals the scale of events observed at particular IP addresses by my data provider. VPNs are 2.8 to 4.6 times larger than their corresponding non-VPN counterparts, and mobile networks are an order of magnitude larger.

Using these features, we employ a 70/30 split of our training and test dataset. The MCC score for this prediction is 89.3%. For most classes, our ability to detect them is above 90%, measured by the high sensitivity in the panel. This is our preferred measure because it measures the fraction of those correctly classified. The most common mistake made by the algorithm is that it classifies an IP as a business when it is in fact a business VPN—the implication of this mistake is that we will underestimate working from home to some degree.

Addressing Potential Concerns with Classifications While the classification performance appears reasonable, the labeled data come from the U.S. There is good reason to believe IPs are assigned differently in different regions. IP address assignment is controlled by one of five network operators, one for each continent. IP addresses tend to be assigned differently based on the ratio of IPs to population, the share of IPs used for different purposes, and network-operator specific practices. In the U.S., it is common for a household to have its own IP address. In some countries, IPs may be shared by small towns. Given this data limitation, while we observe internet activity of firms worldwide, we restrict our study to firms in the United States.

Another issue might be biases contributed by VPNs. From our perspective, someone on a VPN is working remotely. There are edge cases where this is not the case (such as certain government work where we use VPNs in the office), but this will bias down our remote work measure. The broader question is whether VPNs can be easily identified and whether they matter. First, VPNs do not always tunnel traffic. VPNs create latency as one must bounce traffic to another server, and bandwidth could be costly (for example, some VPN providers charge in proportion to bandwidth). They may also want to protect the privacy of the user. Thus companies have an incentive to "leak" traffic that is not strictly necessary. Since publishers we study are public content, they may fall into this category. Suppose further we identify VPN IPs, there are two additional scenarios to consider. First, for some VPNs, the VPN traffic has its own IP address and that VPN IP address is just the address that will be observed by the publisher. Second, the VPN is simply a gateway and the actual traffic is routed to another IP. This is called "network address translation" (NAT). We would only be concerned if the resulting IP address is indistinguishable from the office IP. This might happen in a university for example, because a VPN might be used to access secure resources that are white-listed for a particular IP (such as a library resource). We are unaware of any statistics about the relative frequency of these scenarios.

On the issue of VPNs, we do not foresee large concerns for drawing inferences. First, VPN usage spikes in the beginning of the pandemic, but subsides quickly after the first few months. Firms adjust to perhaps tunneling only specific traffic necessary to VPNs or offer other secure access solutions that make it easy to operate without a VPN. Second, half of the traffic we observe in our dataset pertain to mobile devices. We conjecture that a work device such as a laptop is more likely to be VPN-connected but mobile devices are less likely. Third, our estimates suggest that the VPNs we can observe are relatively rare in the data. These are VPNs which have their own IP, and the traffic that comes from them is not translated via NAT into the office IP. NAT VPNs would have to be at least 10–50 times more common in the data than non-NAT VPNs to be economically significant. Finally, if we cannot detect remote work, this serves as attenuation bias which simply works against being able to observe within-firm variation in remote work.

A.2 Appendix: Estimating and Validating Commute Distance

We construct our measure of commute distance based on pre-pandemic information (internet activity throughout 2019). First, profiles are browser sessions that remain persistent. For a profile associated with a firm (either through a consistent recurrence on a company-associated IP, or a third-party profile database) we can observe every IP address the profile accesses an affiliated member website from. We observe the approximate latitude and longitude of the user from the IP when they access the internet. We estimate commute distance as the distance between the median latitude and longitude of a user during the weekday during work hours versus the median location (latitude and longitude) during the weekend. We compute the haversine distance at the user-level. We exclude employees whose work-time and non-work-time locations are exactly the same (i.e., already working from home). For each firm, we then drop the highest and lowest quartile of distances and compute the average haversine distance of the middle two quartiles. We z-score the measure within zipcode to remove differences in transportation or living costs across geographies.

Our measure of commute distance exhibits a positive correlation with measures from SafeGraph data, which is widely used in the literature to study human mobility patterns. SafeGraph's mobility dataset is based on precise coordinates provided by a phone's global positioning system. SafeGraph's Monthly Patterns dataset describes visitations to establishments in the United States, and has a set of data which describes the visitors to an establishment based on their common day-time location and common-night time location. To calculate commute distance, we first calculate the distance between the establishment location and the common night time location of the visitors, which is often used in this industry as an approximation of home location. We then match the establishment, where possible, to parent company. We merge these two datasets to allow us to compare reading-implied commute distance versus Safegraph commute distance, when SafeGraph data is available. While Safegraph's data is inclusive of both retail visitors and employees, and our dataset only consider employees, we would still expect there to be a positive correlation between these measures. As expected, Figure A.7 shows a strongly positive relationship between our commute distance measure and commute distance inferred from SafeGraph. Our metric has the advantage of (1) being available for many firms in our sample, and (2) being able to distinguish visitors from employees of a firm. The drawback is that our location data is less precise.

A.3 Appendix: Identifying Remote Work from Job Postings

To benchmark our RW measure against the remote jobs measured in Hansen et al. (2023), we obtain job postings data from Revelio Labs, which is sourced from Indeed.com. In order to save on computational costs, we operate on a random subsample of over 10 million job postings from 2020 to 2021. We use GPT4-o to label 1000 sample job postings into four categories: "no mention of remote work", "full time remote work", "hybrid remote work" and "full time in office". We then use this to fine-tune a classifier using transfer learning on this set of postings.

We use the Setfit model due for two reasons. First, it is based on *Sentence Transformers*, which is designed to encode meanings of sentences rather than tokens (like standard transformers models). This means it can operate on multiple sentences, which may be appropriate for job postings - and in fact has been applied in industry practice for job postings and resumes.[34] Second, it is designed to be cost-effective on CPU. Since we classify over 10 million job postings, inference cost must be efficient in order to be feasible on our limited hardware. We train the classifier using default parameters and achieve an overall accuracy of over 80%. This is good accuracy, though not perfect. We restrict to firm with at least 10 job postings over 2020 and 2021 to average out idiosyncratic classification errors.

For each firm with at least 10 job postings over 2020 and 2021, we compute the average share of job postings that are remote. We code fully remote as 1, hybrid remote as 1/2, and full in-person or no mention of remote as 0. We the collapse this to the firm level to measure a firm's share of remote job postings in 2020 and 2021, and correlate this with our remote work measure. Figure 2 shows a strong positive relationship between the fraction of remote job postings by a firm over 2020 and 2021 and our firm-level remote work measure RW_{post}.

There are a number of reasons this elasticity is lower than 1. Both have uncorrelated measurement errors. Job postings do not capture all jobs, only open vacancies. As such, the stock of jobs that were already filled during 2020-2021 will not be captured. Job postings also only reflect intended work arrangement at the time of hiring, which may be different from subsequently realized remote arrangement. Additionally, while we code hybrid as half remote, in reality, firms have discretion over the actual fraction of time workers are allowed to work from home. Relative to the job posting-based measure, our Internet-based RW measure has several advantages: 1) It does not condition on firms with open job postings, which are likely firms with growing labor demand; 2) It is able to capture all jobs, not just jobs with open vacancies; 3) It captures realized remote work at any point in time rather than remote work intended at the time of job posting; 4) It captures finer variation in the fraction of remote work time that is not usually spelled out in job postings.

[34]For example, see https://www.joveo.com/blog/using-sentencebert-to-generate-job-embeddings-for-applications-at-joveo/ retrieved January 2025