Incentivizing Innovation in Open Source: Evidence from the GitHub Sponsors Program*

Annamaria Conti[1], Vansh Gupta[2], Jorge Guzman[3], and Maria Roche[4]

[1] *IE University,* annamaria.conti@ie.edu
[2] *Cornell University & Charles River Associates,* vg224@cornell.edu
[3] *Columbia University & NBER,* jorge.guzman@columbia.edu
[4] *Harvard University,* mroche@hbs.edu

Abstract

Open source is key to innovation, but we know little about how to incentivize it. In this paper, we examine the impact of a program providing monetary incentives to motivate innovators to contribute to open source. The Sponsors program was introduced by GitHub in May 2019 and enabled organizations and individuals alike to reward developers for their open source work on the platform. To study this program, we collect fine-grained data on about 100,000 GitHub users, their activities, and sponsorship events. Using a difference-in-differences approach, we document two main effects. The first is that developers who opted into the program, which does not entail receiving a financial reward, increased their output after the program's launch. The second is that the actual receipt of sponsorship has a long-lasting negative effect on innovation, as measured by new repository creation, regardless of the amount of money received. We estimate a similar decline in other community-oriented tasks, but not in coding effort. While the program's net effect on users' innovative output appears to be positive, our study shows that receiving an extrinsic reward may crowd out developers' intrinsic motivation, diverting their effort away from community and service-oriented activities on open source.

JEL Codes: O3, O31, O36, J24, L86

Keywords: *Open Source*, *Innovation*, *Incentives*, *Financial Rewards*, *Crowding-Out*

*Authors listed in alphabetical order. Many thanks to Shane Greenstein, Frank Nagle, Rob Seamans, and Dan Wang for advice on earlier drafts. Thank you to Jessica Lord for providing deep insight into the program. This manuscript benefited from many helpful comments provided at the AoM, Barcelona School of Economics Summer Forum, the Open Source Workshop, and the NBER Productivity Seminar. Annamaria Conti acknowledges funding from the Swiss National Science Foundation (Project ID: 100013_188998). Maria Roche acknowledges funding from the Harvard Business School Division of Research and Faculty Development. Jorge Guzman acknowledges the support of the Columbia Business School Dean's Summer Fellows program. The views expressed herein are the views and opinions of the authors and do not reflect or represent the views of Charles River Associates or any of the organizations with which the authors are affiliated.

1 Introduction

The role of financial incentives and the provision of these incentives has long been an area of study in innovation research (Arrow, 1963; Boudreau et al., 2011; Cohen and Sauermann, 2007). Ranging from patents, prizes, procurement, and grants, studies have focused on how different features of these incentive regimes may promote innovation (Miric et al., 2019; Wright, 1983). The takeaway of these studies is that the effectiveness of each tool depends fundamentally on appropriability conditions and the nature of idea search (Azoulay and Li, 2020; Moser, 2012).

A parallel line of research emphasizes that open source software has become crucial to innovation (Conti et al., 2021; Nagle, 2019). For example, let us consider the rise of machine learning, which several innovation scholars portray as a fundamental general-purpose technology set to shape productivity and growth for years to come (Agrawal et al., 2018, 2019; Athey, 2018). Unlike many past breakthrough technologies, like radar, semiconductors, or gene-editing, the adoption and advancement of core machine learning technologies relies to a large extent on open source tools, such as Python libraries. In tandem with the rise in the importance of open source for machine learning, we have also come to observe a decline in the value of information technology (IT) patents.[1] Taken together, this may be an early indication for more structural changes in the sourcing for and the production of innovation, at least as it pertains to information technology.

Traditional incentives do not perform well in open source environments for several

[1]This trend has become most apparent since landmark U.S. Supreme Court case rulings that limit their extent of use. For more, please see *Alice Corp. v. CLS Bank International*, for example (Acikalin et al., 2022).

reasons. First, there are important appropriability concerns given that, at its core, open source entails making original code freely available for others to use and build upon. Second, idea search cannot be neatly categorized as either open- or close-ended. Third, the attribution of priority is complicated because open source innovation relies on several small contributions to projects often taking place over long periods of time.

The current understanding of innovation incentives in open source environments appears, if anything, incomplete. Existing work has emphasized the role of intrinsic motivations (Lakhani and Wolf, 2003; Lerner and Tirole, 2002; Nagle et al., 2020), while downplaying pecuniary incentives (Nagle et al., 2020). Yet, financially supported users might improve their productivity and output by, for example, acquiring tools and knowledge they would not have acquired absent financial incentives. At the same time, monetary incentives might crowd out intrinsic motivation (Bénabou and Tirole, 2003) in situations where there is asymmetric information regarding the nature of a financially supported task or developer capabilities. In short, while it is clear that open source innovations provide fundamental contributions to an economy's productivity and growth, there appears to be little guidance on how to incentivize innovative output originating from it.

This paper provides the first evaluation of the impact of a financial incentive regime on innovative output in open source. We examine this by studying a program targeted at motivating innovators to develop open source contributions: the GitHub Sponsors program launched in 2019. This program enabled the financial "sponsorship" of open source developers through the platform, allowing organizations and individuals alike to reward developers financially for their work done on open source software through GitHub. Using data on all 14,892 sponsorable[2] GitHub users from across the globe (as of June 2022) and a corresponding group of 87,310 randomly selected non-sponsorable users, we collect detailed records of each individual's monthly activity on the platform, such as the repositories created and forked, issues opened, and commits, documenting two main relationships regarding the effectiveness of the financial program in incentivizing open source innovation.

First, implementing inverse probability weighting and a difference-in-differences approach, we estimate the extent to which the introduction of the GitHub Sponsors program that allowed some individuals to *become sponsorable* – i.e., to become available to receive a sponsorship – influenced their open source activities. Absent pre-trends, we find that the introduction of the Sponsors program led to an increase in innovation (i.e., new repositories), community-based activities (i.e., issues), and effort (i.e., commits) by over 50%, where both the high and the low end of the quality distribution experienced increases.

Second, we examine the impact of *being sponsored* – i.e., receiving actual financial support – since it is possible that some contributors may put in additional effort because they can now reap financial benefits from their work. For example, monetary compensation may enable developers to divert more of their time to new projects or invest in productivity-enhancing hardware or software. To estimate the relationship of receiving financial support on activity on GitHub, we use the Poisson quasi-MLE two-way Mundlak regression model by Wooldridge (2021), which allows for treatment

[2]Sponsorable users are users who can be sponsored by any other company or individual on GitHub. A user opts into being "sponsorable" by a simple process on GitHub, and only such "sponsorable" users can be sponsored on the GitHub platform.

effect heterogeneity depending on the length of time a user has been treated since the first sponsorship. We apply this model to a homogeneous sample of users who have revealed their willingness to be sponsored. In contrast with the positive effect of the program launch, we document that the actual receipt of a sponsorship *reduces* the level of innovative and community-based activities, although not the frequency of code commits.

To dig deeper into the role of the amount of money rewarded, we make use of a policy implemented by GitHub that ended on January 1, 2020. Namely, to stimulate uptake of the Sponsors program, GitHub pledged to match the amount developers raised from sponsors one to one, up to a total of $5,000 per developer for their first 12 months of being sponsorable (GitHub, 2023). At the end of the 12 months, the amount of money a sponsored developer received for contributions would be essentially halved. Examining the relationship between time to and from the end of the money-matching policy, we fail to detect any relationship at conventional levels. Connecting the totality of these results to our earlier findings suggests that while users invest effort to become visible to and attract potential sponsors, once a monetary incentive is received, the non-monetary, intrinsic motivation to produce new tools for the open source community is crowded out.

Together, our results contribute to the study of innovation incentives in the following two ways. First, because of the distinct nature of the program we study, we are able to focus directly on the way innovation production is influenced by a 'regime change' in the possibility of sponsorship programs, particularly those that focus on unrestricted support for the individual innovator. While the bulk of work on innovator incentives focuses on stimulating innovation for specific projects or goals, such as through procurement, patents, and prizes (Azoulay and Li, 2020; Moser, 2012; Wright, 1983), we instead focus on a program that is targeted at individuals at the higher end of the productivity distribution to work on projects of their own choice on the platform.

Second, we provide novel evidence for the effects of monetary incentives on the work output of open source. Innovation scholars have long been keen to understand what drives investment in open source. For open source contributors, prior work has highlighted a mix of altruism, intrinsic utility, and career concerns (Boudreau and Jeppesen, 2015; Lakhani and Wolf, 2003; Lerner and Tirole, 2002). For firms, investment in open source is often associated with developing absorptive capacity capabilities (Nagle, 2018). Our study provides evidence that the introduction of new regimes promoting financial support for developers can induce an initial boost in productivity, similar to the boost assistant professors and students would experience before obtaining tenure and being admitted into selective graduate programs, respectively. However, once developers receive actual support, the financial incentives crowd out intrinsic ones with long-lasting negative effects on innovation. Therefore, our results urge caution for platform owners. By providing monetary compensation, the sponsor may, for example, unintentionally, reveal that they are not confident the developer will be motivated enough to exert effort in the absence of added incentives signaling to the developer that the task at hand may be difficult and less enjoyable than the developer perceived it to be (Bénabou and Tirole, 2003).

In conclusion, our results highlight a fundamental difference between open source and traditional innovation incentive mechanisms. Therefore, they serve as a clear call for a more systematic analysis of open source innovation incentives in light of the importance of open source for the innovation ecosystem.

The remainder of this paper is structured as follows. In Section II we discuss related literature, in Section III we describe the context and the data. Sections IV and V present the empirical specification and the results, respectively. In Section VI we discuss potential interpretations of our findings as well as contributions to the literature and conclude.

2 Incentivizing Innovators

Although innovation is a fundamental driver of economic growth, the innovation process is fraught with market failures, which implies that innovative effort is under-provided (Bryan and Williams, 2021). Given these market failures, policymakers and scholars have devoted much attention over the past decades to understanding how best to incentivize innovation. Ranging from patents, prizes, procurement, and grants, studies have focused on how different features of these regimes may promote innovation (Azoulay and Lerner, 2013; Azoulay and Li, 2020; Moser, 2012; Wright, 1983). Typically, the decision to implement a specific incentive scheme depends on weighing factors such as appropriability conditions and the nature of the idea search. Though effective in certain contexts, these regimes may not translate to settings where it is infeasible or undesirable to enforce intellectual property rights and specify ex-ante the goals that participants should attain, or what the end product should look like.

Open source, a software development model where a body of original material is made publicly available for others to use, under certain conditions, is one such context. Open source software advances innovation by having users solve problems valuable to the community. Today, over 90% of all Fortune-500 companies use open source products (Github, 2022), and their role is increasingly relevant in the development of the software industry.

Given differences in the nature of open source, producing and maintaining open source innovation may be distinct from traditional intellectual property rights. For example, in open source, idea search can be both open-ended or have a known-end falling between existing incentive regimes that are designed to promote the one or the other. Moreover, open source contributions, which are typically smaller than those considered under other regimes, require regular effort by many participants and the nature of software innovation is inherently cumulative. Such a cumulative nature makes assigning intellectual property rights difficult, if not entirely infeasible. Furthermore, the outcome of participants' efforts is, by definition, public, reducing the ability of innovators to enforce monopoly power for financial returns from their effort ex-post.

What, then, motivates open source developers to contribute to open source projects? Innovation research has documented survey evidence that developers derive an "intrinsic joy" from participating in open source and they prefer such involvement rather than performing routine tasks set by an employer (Lakhani and von Hippel, 2003; Nagle et al., 2020). Indeed, users on open source platforms may resemble scientists who are driven by intrinsic motivation and non-pecuniary incentives such as the joy of puzzle solving, peer recognition, and interaction with other scientific community members (Dasgupta and David, 1994; Merton, 1973; Stern, 2004). Under this perspective, financial incentives are unlikely to be useful in motivating developers to contribute to the open source.

While the intrinsic motivation / non-pecuniary incentive perspective has gained traction, there is also a literature stream that advocates the importance of financial incentives. This stream has highlighted that whereas intrinsic motivation and non-pecuniary incentives in open source may be motivating factors, financial and career concerns may also play an important role. For one, participation in open source may favor internal career promotions provided that open source knowledge increases performance in paid work (Lerner and Tirole, 2002; Nagle, 2018). For another, open source could lead to future job offers, self-employment opportunities, and access to venture capital. Lakhani and Wolf (2003), Boudreau and Jeppesen (2015), Xu et al. (2020), Conti et al. (2021), Jeppesen and Frederiksen (2006) find some evidence consistent with these conjectures. In general, financial compensations could allow users to undertake projects that they would not have been otherwise able to undertake had they not used the payment received to buy needed hardware or software, for example. Moreover, it is possible that some contributors may put in more effort because they can now reap financial benefits from such work, enough even to seek self-employment.

Although financial incentives may boost innovation activities in open source, prior work examining extrinsic and intrinsic motivation highlights the potential for crowding out of intrinsic motivation in the long run (Bénabou and Tirole, 2003; Frey and Oberholzer-Gee, 1997). By providing monetary compensation, the sponsor may, unintentionally, reveal that they are not confident the developer will be motivated enough to exert effort in the absence of added incentives. Additionally, such compensation may reveal that the task at hand is difficult and less enjoyable than the developer has perceived it to be. If this is the case, we might observe a decline in the innovative activity of those who receive a financial reward.

Given these arguments, it is *a priori* unclear whether direct financial rewards stimulate, have no impact on the net, or even lead to a decrease in innovative contributions to open source. In what follows, we will provide empirical evidence on a specific monetary incentive program that was put in place by the largest open source platform at the time of this study - GitHub. The program in question was initiated by GitHub in 2019, which enabled financial sponsorship of open source developers. The Sponsors program enabled organizations and individuals alike to pay developers for their knowledge work done on the open source platform, providing a useful context for us to examine to what extent financial rewards may stimulate open source innovation.

3 Empirical Context and Data

Our goal is to assess the effect of the GitHub Sponsors program on various aspects of user behavior on the platform, and to evaluate whether financial rewards may impact user participation in open source and, ultimately, innovation.

3.1 The GitHub Sponsors Program

The Sponsors program was launched by GitHub in May 2019 to help support open source developers financially (Zuegel, 2019). The goal of the program's founder was to give open source developers another option for making a living, and to provide legitimacy to the idea of developers asking for support from the community that depends on the infrastructure they built (Zuegel, 2019). The Sponsors program is somewhat similar to Patreon, a membership platform that provides business tools for content creators to be financially rewarded by so-called “patrons”. In contrast with Patreon, the GitHub Sponsors program focuses specifically on open source contributions. Moreover, GitHub does not charge any fees for sponsorships from personal accounts, that way, 100% of the sponsorships go to the sponsored developer. GitHub does, however, charge a 10% fee for sponsorships from organizations.

GitHub first launched a beta version of the Sponsors program in May 2019. During the beta period, GitHub predominantly prioritized developers who had open source software projects that many people depend on—that is, developers already prominent in the space. It opened the program to all users in September 2019, once it managed to develop an automated approach for account approval. The requirements to become a sponsorable user are minimal: the application requires that a requesting developer is over the age of 13, has agreed to GitHub's terms of acceptable use, and is a tax resident in one of the countries that GitHub has rolled out the Sponsors program to. Sponsorable users are granted flexibility in setting up their accounts and the terms of receiving sponsorships. Most sponsorable users set pre-defined tiers describing the benefits the sponsors may receive by committing a certain amount of money. For example, sponsors may receive a mention in the documentation of the sponsorable's most popular repository, priority support for any issues that a sponsor wants to be fixed in a repository, or even a 1-hour consulting session on anything the sponsor may request the sponsorable's advice on. These tiers allow the sponsorables to set expectations for their sponsors, as well as provide them with the potential to develop long-term relationships with their sponsors, who may be individuals or organizations.

To promote the adoption of the program by users, GitHub introduced a matching program, of up to $5,000 of any user's sponsorship earnings for the duration of their first year in the program (Zuegel, 2019). This incentive effectively allowed sponsorables to double their earnings in their first year in the program. This matching program was introduced for a limited period and stopped accepting applications on January 1, 2020.

3.2 Data Construction

To assess the effect of the Sponsors program, we compile a novel dataset using a combination of the GitHub Search API (*GitHub Search API*, 2022) and the GitHub Archive database (Grigorik, 2022). Together, these data provide a detailed window into the behavior and activity of GitHub users.

We begin by identifying those users who opt into the GitHub Sponsors program and collect their usernames employing the GitHub Search API. Overall, there are 9,955 so-called "sponsorable" users. These users are distributed worldwide, with a majority being in North America and Europe. We next use GitHub GraphQL API to create an unbalanced panel of these users, their sponsorship events, and activities until June 1, 2022. It is noteworthy that the GitHub GraphQL API only provides publicly available information. While private sponsorship events are publicly available information and are recorded with precise time stamps, information about the sponsors—including the sponsors' country of origin and whether they are individuals or organizations—is only available if the sponsors decide to disclose their identity. Our panel further includes information about the repositories created by the users, such as the repository names, creation dates, licenses, number of stargazers[3] and number of forks. We additionally identify whether a user created a repository from scratch or forked an existing repository. We collect additional data on users' issues created and commits using GitHub Archive. This program continuously scrapes and stores all activity conducted on GitHub in the public domain.

3.3 Summary Statistics

Of the 9,995 individual users who became sponsorables between May 2019 (the start date of the program) and June 2022 (the date at which we retrieved the GitHub information), 32.3% ever got sponsored as of June 2022. Even amongst those who did get sponsored, there exists a large amount of heterogeneity in how frequently they are sponsored. As reported in Figure 1, there has been a steady increase since the introduction of the Sponsors program in the number of first-time sponsorships over time, especially after the termination of the beta period.

In the first part of the empirical analysis, we assess the impact of becoming sponsorable, that is, becoming part of the GitHub Sponsors program, regardless of whether sponsorable users were eventually sponsored. To find a suitable control group for this analysis, we use the GitHub Archive dataset to randomly sample 46,152 users who had at least one commit or pull request in the first three months of 2018, that is, prior to the introduction of the Sponsors program, but did not enroll in the program. For this group, we downloaded the same type of information we retrieved for the

[3]Stargazers are users who bookmark the focal user's repositories.

sponsorable users.

Table 1a reports descriptive statistics distinguishing between sponsorable and non-sponsorable users. We present descriptives for the total number of repositories, commits, and issues generated during the pre-sample period, which starts in September 2018 and ends in December 2018. A repository (or 'repo') is a compilation of files for a specific project, saved in GitHub in a dedicated folder. Such repositories can be created from scratch (hence are referred to as 'original') or can be created by copying someone else's repository and modifying it through a process called 'forking' (hence referred to as 'forked'). We view repositories as the closest equivalent to innovation since creating such repositories requires a considerable amount of knowledge, skills, and coding effort, and they predominately are stand-alone tools (e.g., freeCodeCamp, Awesome, Build Your Own X, TensorFlow). Commits refer to a single batch of changes users make to files in a repository. As such, they are a suitable proxy for the effort users exert to modify their own code. Issues refer to the practice of documenting problems with the code in a repository. The documentation of problems through issues allows multiple people to collaborate easily on open-source repositories. Therefore, issues can be considered proxies for the contributions users provide to open source.

As reported in Table 1a, sponsorable users create 0.47 more repositories than other users. The difference between user types is larger when we examine forked repositories than when we consider original repositories, although it is highly significant in both instances. The table additionally shows that sponsorable users generate more commits and issues than other users.

Even though sponsorable users appear to be more prolific than the other users, we observe in Figure 2 that the *growth* pattern in the number of repositories created each month is the same across sponsorable and non-sponsorable users before the launch of the Sponsors program. It diverges afterwards, especially in the post-beta period.

In the second part of the empirical analysis, we assess the impact of obtaining a first sponsorship on subsequent user activities. As we detail in the next section, we restrict this analysis to sponsorable users and compare each sponsored user with users that were never sponsored. In Table 1b, we report descriptive statistics distinguishing sponsorable users according to whether they had been sponsored as of June 2022. As shown, sponsored users are more active on GitHub than other sponsorables across the full spectrum of GitHub activities during the pre-sample period. However, Figure 3 shows that the growth pattern in the number of repositories created each month is the same across sponsored and unsponsored users *both* before and after sponsored users start being sponsored.

4 Empirical Methodology

Our paper plans to address the following two empirical questions: 1) whether *being admitted* to the GitHub Sponsors program and, thus, becoming sponsorable, affect the innovative output of the program members, and 2) whether *obtaining* an *actual sponsorship* has any impact on innovative output.

4.1 The Impact of the Launch of the Sponsors Program on Sponsorables

To assess the effect of becoming "financially rewardable" through the GitHub Sponsors program on users' innovative activities in open source, we compare the behavior of users who eventually became sponsorable, before and after the launch of the Sponsors program, to that of similar users who did not participate in the Sponsors program. As we do not observe the timing at which individuals became sponsorable, we treat this characteristic as a time-invariant user feature. This analysis aims to assess whether users might increase their innovation efforts, being incentivized by the possibility of receiving a financial reward.

As we showed in Table 1a, sponsorable users differ substantially from the other users. They create more repositories and issues and submit more commits. In principle, these user discrepancies should not be a problem for our difference-in-differences approach as long as the parallel trends hold. However, to make the two samples more comparable and address the possibility that these intrinsic differences may bias our main effects, we implement inverse probability weighting (IPW) (Imbens, 2000). We begin by estimating, for each observation, the probability of receiving treatment, that is, becoming sponsorable, through a probit model. Specifically, we model the probability of being treated as a function of measures for user activities on GitHub observed in the pre-sample period, between September and December 2018, before the program was announced. These measures are: the total number of repositories, commits, and issues created, the natural logarithms of these counts, and the square of the logarithms.

Having estimated the probit model, we predict each user's probability p of becoming sponsorable. The users that became sponsorable are then weighted by $1/p$, whereas those that did not participate in the Sponsors program are weighted by $1/(1-p)$. Therefore, each observation is given as weight the inverse of the probability of their status. As a result of applying this approach, treated users that resemble the control group are given more weight, and controls that resemble the treatment group are also given more weight. Table A1 shows that this procedure dramatically reduces the differences in pre-sample characteristics between sponsorable users and other users.

Using the generated weights, we estimate the following difference-in-differences Poisson model at the level of user i observed in year-month t:

$$E(Y_{it}|X_{it}) = \gamma_i \exp(\beta \cdot Sponsorable_i \cdot I[t > May\ 2019] + \tau_t), \tag{1}$$

where Y_{it} is the number of repositories, issues, or commits, created by i at time t. $Sponsorable_i$ is an indicator equal to 1 if user i participates in the Sponsors program and 0 if not. The coefficient of interest in this model is β. It represents the change in outcome after the Sponsors program was introduced for treated users relative to untreated users. The γ_i are user fixed effects that absorb fixed differences across users, while τ_t are year-month fixed effects. In alternative specifications, we distinguish repositories according to whether they are original or built on other users' input and by their quality. These measures proxy for the innovative output produced by user i on open source. $I[t > May\ 2019]$ is an indicator equal to 1 for the period after May 2019, when the GitHub Sponsors program was launched, and 0 otherwise.

4.2 The Effect of Being Sponsored

To estimate the impact of actually *being sponsored* on user open source innovation activities, we estimate a variant of a standard two-way fixed effect (TWFE) Poisson model, exploiting variation in the timing of receiving a first sponsorship as a source of identification. Recently, TWFE models with unit and time fixed effects have come under considerable scrutiny as they may deliver inconsistent estimates if treatment effects are heterogeneous across groups or time (Callaway and Sant'Anna, 2021; De Chaisemartin and d'Haultfoeuille, 2020; Goodman-Bacon, 2021). To address this problem, we follow Wooldridge (2021) and estimate a Poisson quasi-MLE two-way Mundlak regression model, which allows for substantial treatment effect heterogeneity. Specifically, for user i, observed in year-month t, belonging to treatment cohort r, we estimate:

$$E(Y_{it}|d_{iq},...,d_{iT},\gamma_i) = \gamma_i \exp[\sum_{s=2}^{T}\theta_s fs_t + \sum_{r=q}^{T}\sum_{s=r}^{T}\tau_{rs}(Sponsored_{it}d_{ir}fs_t)], \qquad (2)$$

where Y_{it} is, alternatively, the number of repositories, issues, and commits created by user i. As before, γ_i are user fixed effects. The fs_t are time dummies for year-month t, with $fs_t = 1$ if $s = t$ and $fs_t = 0$ if $s \neq t$. $Sponsored_{it}$ is a (0/1) indicator that becomes 1 after user i receives a first sponsorship. The $d_{iq},...,d_{iT}$ are treatment cohort dummies identifying when a user was first sponsored. The expression $Sponsored_{it}d_{ir}fs_t$ corresponds to interacting treatment cohort dummies with time dummies for those periods where a cohort of users is sponsored for the first time. Therefore, τ_{rs} is the average treatment effect on the treated (ATT) for cohort r in year-month t (conditional on $d_r = 1$) for $t = r,...,T, r = q,...,T$. This functional form – which has been shown to deliver consistent estimates – allows for substantial treatment effect heterogeneity, depending on the length of time a user has been treated since the first sponsorship. We apply this model to the sample of sponsorable users defined in Table 1b. In fact, this is the category of users at risk of being sponsored, given that they have expressed explicit interest in receiving financial compensation by becoming sponsorables.

5 Results

5.1 The Impact of the Launch of the Sponsors Program on Sponsorables

We begin by discussing the effect of becoming sponsorable as a result of the launch of the Sponsors program on the creation of new repositories by users. The results from estimating the Poisson model described by Eq. (1) are reported in Table 2, where we cluster standard errors at the user level. The relevant sample consists of sponsorable and non-sponsorable users, where each observation is weighted by the inverse propensity score of becoming sponsorable.

In column 1, we show that the introduction of the Sponsors program leads to an increase of 0.43 in the sponsorables' hazard of creating a new repository each month. That is, after GitHub launched the Sponsors program, sponsorable users created 54% more repositories relative to non-sponsorable users each month. To the extent that creating a repository implies inventing a new open source project, this effect suggests that sponsorable users generated more innovative output after the launch of the Sponsors program. In column 2, we distinguish between the beta period and the period after. Here, we show that after the introduction of the Sponsors program, the hazard of creating a new repository increased by 0.24 during the beta period and by an additional 0.26 afterwards. These figures imply that during the beta period, sponsorables generated 27% more repositories than non-sponsorable users. After the end of the Sponsors' beta period, sponsorable users generated an additional 30% more repositories than non-sponsorables, relative to the beta period. These results suggest that the spur in innovative output we observe after the launch of the Sponsors program is not limited to the beta period.

The effect of the launch of the Sponsors program on the number of original repositories (column 3) is slightly larger than the effect on the number of forked repositories (column 5). Distinguishing between the pre- and post-beta period, we show that sponsorable users generated more original repositories during the beta period than after (column 4). However, the effects are flipped when we consider the number of forked repositories as an outcome (column 6).

Figure 4 reports the results from a dynamic version of Eq. (1) that replaces the $Sponsorable_i \cdot I[t > May\ 2019]$ term with time indicators for each year-month to/from the launch of the Sponsors program. The reported coefficients represent the additional proportion of new repositories created each month by sponsorable users relative to non-sponsorables. In Panel A, we examine the total number of new repositories created; in Panel B, we use the number of forked repositories; while in Panel C, we assess the number of original repositories. We highlight three main patterns. First, we observe no pre-trends regardless of the outcome examined, suggesting that new repositories by sponsorable and non-sponsorable users would have followed a similar trend in the absence of the launch of the Sponsors program. Second, we observe an uptick in the number of new repositories after the conclusion of the beta period. Third, the effects of the Sponsors program remain positive and significant up to twenty months after its launch.

In Table 3, we delve deeper into the quality of repositories sponsorable users create as a result of the launch of the Sponsors program. Column 1 replicates our baseline result for the total number of repositories generated, equivalent to the results in column 2 of Table 3. In column 2, we examine only the high-quality repositories, proxied by those in the 95$^{\text{th}}$ percentile of the distribution for the number of forks received as of December 2022. As shown, the hazard of generating high-quality repositories increases by 0.20 for sponsorable users during the beta period and by an additional 0.29 after the end of the beta period. These effects are comparable to those reported in column 1, implying that sponsorable users create 22% more high-quality repositories than non-sponsorable users during the Sponsors' beta period, and an additional 34% high-quality repositories after the beta period ends. We find similar results in column 3 when we adopt an alternative definition of repository quality by focusing on those in the 95$^{\text{th}}$ percentile of the distribution for the number of stars obtained as of December 2022.

Columns 4 and 5 display the results for repositories of lower quality. As shown, sponsorable users also increase the production of repositories with no forks or stars, although the increment in the production rate is considerably lower in the post-beta period relative to the production rate of high-quality repositories. Overall, these results suggest that the Sponsors program might have succeeded in boosting not only the volume of sponsorables' open source contributions but also their quality.

Extending these findings, in Table 4 and Figure A1, we report the results for the number of new commits and issues created. The displayed effects are consistent with our main findings in Tables 2 and 3 and Figure 4. Sponsorable users produce more commits and issues than non-sponsorable users after the launch of the Sponsors program.

Overall, these results suggest that after the launch of the Sponsors program, sponsorable users increase their participation in open source across the full spectrum of activities and also generate, new, innovative code. One remaining question is how significant the boost in innovative activities is. To address this question, we highlight that the 54% increase in repositories, inferred from column 1 of Table 2, implies that the average sponsorable user would generate a monthly number of repositories equal to 1.24 post-program-launch, up from 0.84. This translates into 15 repositories per year which, multiplied by the 14,892 sponsorable users available as of June 2022, corresponds to 223,380 repositories. This is not a small number, particularly given that repositories created by sponsorable users are more valuable for the open source community.

5.2 The Impact of Being Sponsored

Having shown that the launch of the Sponsors program induced sponsorable users to increase their contributions to the platform as well as their innovative output, the next question we address is whether this incentive is reinforced after sponsorable users obtain their first sponsorship. Specifically, we assess the relationship between receiving a first sponsorship and creating new repositories. To do so, we estimate the Poisson model described by Eq. (2) applied to the sample of sponsorable users. The results are displayed in Table 5. The reported effects correspond to the average marginal effects over treated cohorts and time, where the control group is formed by never-treated users. Standard errors are clustered at the user level.

As reported in column 1, after sponsorable users obtain a first sponsorship, the hazard of generating a new repository each month declines by 0.15, implying that the production of new repositories declines by 16%. The magnitude of the decline is similar whether we examine original repositories (column 2) or forked repositories (column 3). We also show in column 5 that obtaining a first sponsorship has a negative impact on the number of new issues generated but an insignificant effect on the number of new commits (column 4). Combined, these results portray an interesting picture. As we mentioned earlier, commits represent the effort users exert to write and modify their own code and, thus, an input in the production function of users' innovative output. Hence, the non-significant results we detect for the number of commits paired with the negative effects of receiving sponsorship on the number of new repositories and issues suggest that while sponsored users did not reduce effort on writing code, they diverted it away from creating new projects and supporting community activities through issues. The fact that effort does not drop, but rather certain types of activity do, suggests that, for one, individuals do not leave the platform once they receive sponsorship, and for another, that our results are not driven by a tournament-type explanation. Though a feasible alternative explanation, the totality of our results would appear inconsistent with this interpretation since, if indeed a race were to be driving the results, effort as measured by commits should also change in response to receiving sponsorship.

An analysis of the dynamic effects displayed in Figure 5 for the twelve months before and after sponsorables obtain their first sponsorship reveals a nuanced picture. As shown, sponsored users generate more repositories the first month they are sponsored relative to never sponsored users *but* fewer repositories thereafter. This pattern is consistent regardless of whether we examine the total number of repositories

(Panel A), the number of forked repositories (Panel B), or the number of original repositories created (Panel C). We observe a similar pattern in Panel B of Figure 6 where we examine the number of new issues created. Conversely, Panel A of Figure 6 depicts a flat trend for the number of new commits after sponsorable users obtain their first sponsorship. Importantly, because we observe no significant pre-trends, it is unlikely that sponsored user activities on GitHub decline post-sponsorship as a result of reversion to the mean.

Delving deeper into these results, Table 6 shows that the negative effect of obtaining a first sponsorship is especially significant in the case of repositories without forks (column 4). However, the effect remains negative and strong even when top repositories by the number of forks (column 2) and stars received (column 3) are considered. Overall, these results suggest that being granted sponsorship might reveal information regarding the characteristics of remunerated tasks, possibly changing users' perception of the open source contract and crowding out non-pecuniary incentives to contribute innovation to the open source in the longer run. To be clear, the program's net effect on users' innovative output is positive as its launch induced sponsorable users to produce more innovative output. However, receiving an extrinsic reward appears to divert sponsored users' innovative efforts away from community and service-oriented activities on open source.

We next examine an alternative explanation to crowding out, that is, users may decrease their innovative output and contributions to open source because the sponsorship amount they receive is inadequate. A small amount may reveal information to sponsored users that the compensations they could receive are insufficient to make a living from the platform. To assess this possibility, we exploit the fact that when GitHub launched the Sponsors program, it matched up to $5,000 USD of any user's sponsorship earnings for the duration of their first year in the program. If inadequate compensation were the cause of the negative effects of being sponsored, we would expect the detected negative effects to be stronger during the months after the matching by GitHub ends. To verify this, we modify Eq. (2), identifying the period after one year from the first sponsorship as the treatment period. The reported effects in Table 7 are no more negative than the ones displayed in Table 5. These findings are corroborated by Figure 7, which shows that the trends are flat after GitHub's fund matching stops. Overall, these results confirm that obtaining financial support does not boost innovative output on open source and suggest that the possibly limited reward size may not be the prevalent explanation.

6 Conclusion

The role of incentives for innovation has long been the focus of research by innovation scholars. While the literature has devoted much attention to instruments such as patents in contexts where appropriability is feasible, little is known about how to incentivize innovative output in open source, where the process of producing innovative ideas is cumulative and appropriability is unfeasible or undesirable. To date, vis-à-vis other innovation areas such as patented work, open source remains with relatively few monetary incentives for innovators and the concern remains if open source is sustainable or can be sustained for years to come. Yet ensuring the sustainability of the open source model is of crucial importance as innovations originating from open source software environments have become a key input for firms.

In this paper, we examine to what extent the goal of a specific financial program targeted at motivating innovators to develop critical contributions to open source is achieved. To do so, we exploit the introduction of a program initiated by GitHub in 2019, which enabled "sponsorship" of open source developers. We document two main effects regarding the effectiveness of the program.

First, we analyze the impact of the launch of the Sponsors program on sponsorable developers estimating difference-in-differences models with inverse probability weights. Absent noticeable pre-trends, our results suggest that the introduction of the Sponsors program led to an increase in new knowledge and code edits. Second, we examine the potential impact of obtaining sponsorship using a Poisson quasi-MLE two-way Mundlak difference-in-differences regression model. Our results indicate a long-term decline in activities involving repositories, our measure of innovative activity, and community-oriented activity. The amount of commits – our measure for effort – however, remains unchanged. We show that these relationships are not driven by inadequate compensation users may receive. In fact, exploiting a policy implemented by GitHub where GitHub matched the amount of money developers raised from sponsors one-to-one, we fail to detect any significant changes in the effect of GitHub ending the matching on sponsored activity on the platform.

Overall, we interpret these results as evidence that the program led to an increase in contributions in anticipation of receiving funding. However, the actual receipt of sponsorship may have potentially changed the perception of contributing new innovation as a form of community service to a market exchange (Gneezy and

Rustichini, 2000) thereby crowding out intrinsic motivation. While the launch of the Sponsors program might have induced a large initial response from users aspiring to receive sponsorship, the program per se might not have succeeded in keeping users engaged and innovative by means of financial rewards.

Naturally, both our data and approach may be limited. For one, the data we collect only include public data that were partially assembled by a third-party. To ensure the reliability of the data, we perform a host of quality checks. For another, our approach may be restricted in its external validity, given that we focus on a specific open source platform and specific regime change. However, provided that GitHub is the largest host of source code with over 40 million public repositories to date, we believe they have broader implications, especially for the type of innovators we study.

Together, our results contribute to the study of innovation incentives in at least two ways. First, provided the distinct nature of the program, we are able to focus directly on the way innovation production is influenced by a 'regime change'. Instead of focusing on innovator incentives that stimulate innovation for specific projects or goals (Azoulay and Li, 2020; Moser, 2012; Wright, 1983), our context particularly focuses on unrestricted support for the individual innovator. We thereby shift our attention to a program that is targeted at individuals at the high end of the productivity distribution to work on projects on an open source platform.

Second, our study provides novel evidence that the introduction of an incentive program can generate meaningful impetus for contributions and programmer productivity, and turns negative for those who receive financial rewards in the long run.

This result is consistent with, for example, research examining the impact of tenure on academics' output. Findings in this line of work indicate that while academics invest significantly in their research to receive tenure, their output tends to drop following this event (Brogaard et al., 2018). Further studies in this area suggest that such tenure-style regimes also found in civil service, law and accounting firms represent a form of risk-sharing when human capital investments are in question (Ito and Kahn, 1986). Similarly, the connection can be made to the productivity boost students would experience before being admitted into selective graduate programs. Once they are accepted to the programs, their motivation to excel in coursework is often reduced. Though our context may be perceived as somewhat lower stakes, reputation concerns carrying potential long term consequences for coders' careers cannot be ignored (Xu et al., 2020).

Overall, this study provides insight into a type of incentive regime that has remained underexplored in the context of innovation: Sponsorship. Generally, the idea prevails that innovation, which is critical for economic growth, benefits from monetary incentives (Romer, 1994). Existing regimes relying on such incentives, such as patents (Gallini and Scotchmer, 2002), prizes (Galasso and Schankerman, 2018; Murray et al., 2012), grants, and procurement (Azoulay and Li, 2020) have been shown to be effective in certain contexts suggesting that sponsorship may too. However, not all innovators are motivated by profit, but may be driven by, e.g., impact (Cohen et al., 2020; Guzman et al., 2023), and personal needs (von Hippel, 1988). As this paper demonstrates, particularly in the case of open source, where developers appear to be community and service oriented, the inclusion of monetary incentives remains a fragile balancing-act.

References

Acikalin, U. U., Caskurlu, T., Hoberg, G. and Phillips, G. M. (2022), Intellectual property protection lost and competition: An examination using machine learning, Technical report, National Bureau of Economic Research.

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

Agrawal, A., Gans, J. S. and Goldfarb, A. (2019), 'Artificial intelligence: The ambiguous labor market impact of automating prediction', *Journal of Economic Perspectives* **33**(2), 31–50.

Arrow, K. J. (1963), 'Utility and expectation in economic behavior.'.

Athey, S. (2018), The impact of machine learning on economics, *in* 'The economics of artificial intelligence: An agenda', University of Chicago Press, pp. 507–547.

Azoulay, P. and Lerner, J. (2013), 'Technological innovation and organizations', *Handbook of Organizational Economics* pp. 575–603.

Azoulay, P. and Li, D. (2020), *Scientific grant funding*, Vol. 26889, National Bureau of Economic Research Cambridge, MA.

Boudreau, K. J. and Jeppesen, L. B. (2015), 'Unpaid crowd complementors: The platform network effect mirage', *Strategic Management Journal* **36**(12), 1761–1777.

Boudreau, K. J., Lacetera, N. and Lakhani, K. R. (2011), 'Incentives and problem uncertainty in innovation contests: An empirical analysis', *Management science* **57**(5), 843–863.

Brogaard, J., Engelberg, J. and Van Wesep, E. (2018), 'Do economists swing for the fences after tenure?', *Journal of Economic Perspectives* **32**(1), 179–194.

Bryan, K. A. and Williams, H. L. (2021), Innovation: market failures and public policies, *in* 'Handbook of industrial organization', Vol. 5, Elsevier, pp. 281–388.

Bénabou, R. and Tirole, J. (2003), 'Intrinsic and Extrinsic Motivation', *The Review of Economic Studies* **70**(3), 489–520.

Callaway, B. and Sant'Anna, P. H. (2021), 'Difference-in-differences with multiple time periods', *Journal of Econometrics* **225**(2), 200–230.

Cohen, W. M. and Sauermann, H. (2007), 'Schumpeter's prophecy and individual incentives as a driver of innovation', *Perspectives on Innovation* pp. 73–104.

Cohen, W. M., Sauermann, H. and Stephan, P. (2020), 'Not in the job description: The commercial activities of academic scientists and engineers', *Management Science* **66**(9), 4108–4117.

Conti, A., Peukert, C. and Roche, M. P. (2021), 'Beefing it up for your investor? open sourcing and startup funding: Evidence from github', *Open Sourcing and Startup Funding: Evidence from GitHub (August 25, 2021)* .

Dasgupta, P. and David, P. (1994), 'Toward a new economics of science', *Research Policy* **23**(5), 487–521.

De Chaisemartin, C. and d'Haultfoeuille, X. (2020), 'Two-way fixed effects estimators with heterogeneous treatment effects', *American Economic Review* **110**(9), 2964–2996.

Frey, B. S. and Oberholzer-Gee, F. (1997), 'The cost of price incentives: An empirical analysis of motivation crowding- out', *American Economic Review* **87**(4), 746–755.

Galasso, A. and Schankerman, M. (2018), 'Patent rights, innovation, and firm exit', *The RAND Journal of Economics* **49**(1), 64–86.

Gallini, N. and Scotchmer, S. (2002), Intellectual property: When is it the best incentive system?, *in* 'Innovation Policy and the Economy, Volume 2', National Bureau of Economic Research, Inc, pp. 51–78.
URL: *https://EconPapers.repec.org/RePEc:nbr:nberch:10785*

Github (2022), 'Octoverse: The state of open source software 2022'.

GitHub (2023), 'Github sponsors matching fund'.
URL: *https://docs.github.com/en/site-policy/github-terms/github-sponsors-additional-terms22-github-sponsors-matching-fund*

GitHub Search API (2022).
URL: *https://docs.github.com/en/rest/search?apiVersion=2022-11-28*

Gneezy, U. and Rustichini, A. (2000), 'Pay enough or don't pay at all', *Quarterly Journal of Economics* **115**(3), 791–810.
URL: *http://www.jstor.org/stable/2586896*

Goodman-Bacon, A. (2021), 'Difference-in-differences with variation in treatment timing', *Journal of Econometrics* **225**(2), 254–277.

Grigorik, I. (2022), 'Gh archive database'.
URL: *https://www.gharchive.org/*

Guzman, J., Joohyun Oh, J. and Sen, A. (2023), 'Climate change framing and innovator attention: Evidence from an email field experiment', *Proceedings of the National Academy of Sciences* **120**(3), e2213627120.

Imbens, G. W. (2000), 'The role of the propensity score in estimating dose-response functions', *Biometrika* **87**(3), 706–710.

Ito, T. and Kahn, C. (1986), 'Why is there tenure?'.

Jeppesen, L. B. and Frederiksen, L. (2006), 'Why do users contribute to firm-hosted user communities? the case of computer-controlled music instruments', *Organization Science* **17**(1), 45–63.

Lakhani, K. R. and von Hippel, E. (2003), 'How open source software works: "free" user-to-user assistance', *Research Policy* **32**(6), 923–943.

Lakhani, K. R. and Wolf, R. G. (2003), 'Why hackers do what they do:understanding motivation and effort in free/open source software projects', *Open Source Software Projects (Sept. 2003)* .

Lerner, J. and Tirole, J. (2002), 'Some simple economics of open source', *Journal of Industrial Economics* **50**(2), 197–234.

Merton, R. K. (1973), *The sociology of science: Theoretical and empirical investigations*, University of Chicago press.

Miric, M., Boudreau, K. J. and Jeppesen, L. B. (2019), 'Protecting their digital assets: The use of formal & informal appropriability strategies by app developers', *Research Policy* **48**(8), 103738.

Moser, P. (2012), 'Innovation without patents: Evidence from world's fairs', *Journal of Law and Economics* **55**(1), 43–74.

Murray, F., Stern, S., Campbell, G. and MacCormack, A. (2012), 'Grand innovation prizes: A theoretical, normative, and empirical evaluation', *Research Policy* **41**(10), 1779–1792.

Nagle, F. (2018), 'Learning by contributing: Gaining competitive advantage through contribution to crowdsourced public goods', *Organization Science* **29**(4), 569–587.

Nagle, F. (2019), 'Open source software and firm productivity', *Management Science* **65**(3), 1191–1215.

Nagle, F., Wheeler, D. A., Lifshitz-Assaf, H., Ham, H. and Hoffman, J. L. (2020), 'Report on the 2020 foss contributor survey'.

Romer, P. M. (1994), 'The origins of endogenous growth', *Journal of Economic perspectives* **8**(1), 3–22.

Stern, S. (2004), 'Do scientists pay to be scientists?', *Management Science* **50**(6), 835–853.

von Hippel, E. (1988), *The Sources of Innovation*, Oxford University Press.

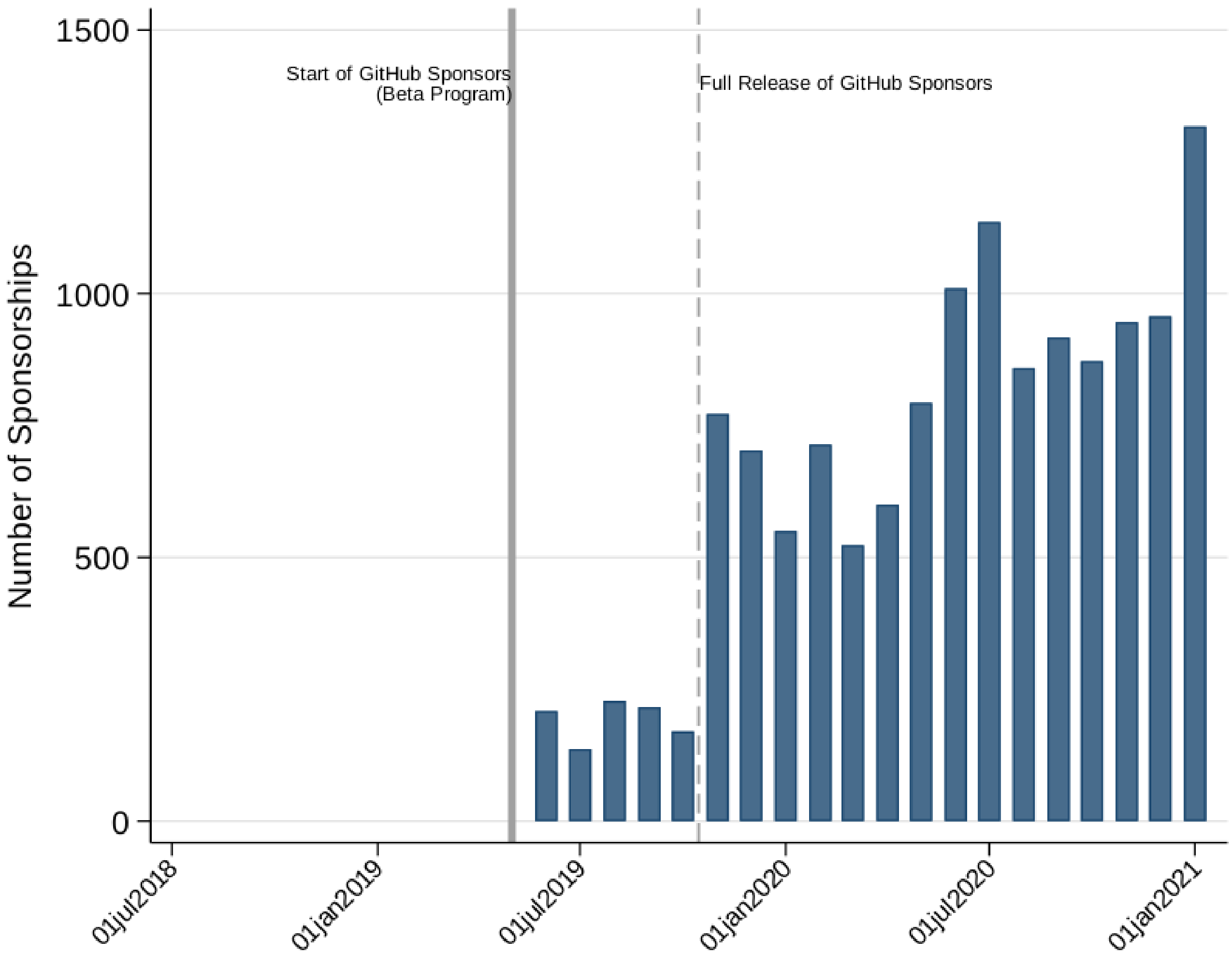
Wooldridge, J. M. (2021), 'Two-way fixed effects, the two-way mundlak regression, and difference-in-differences estimators', *Available at SSRN 3906345* .

Wright, B. D. (1983), 'The economics of invention incentives: Patents, prizes, and research contracts', *American Economic Review* **73**(4), 691–707.

Xu, L., Nian, T. and Cabral, L. (2020), 'What makes geeks tick? a study of stack overflow careers', *Management Science* **66**(2), 587–604.

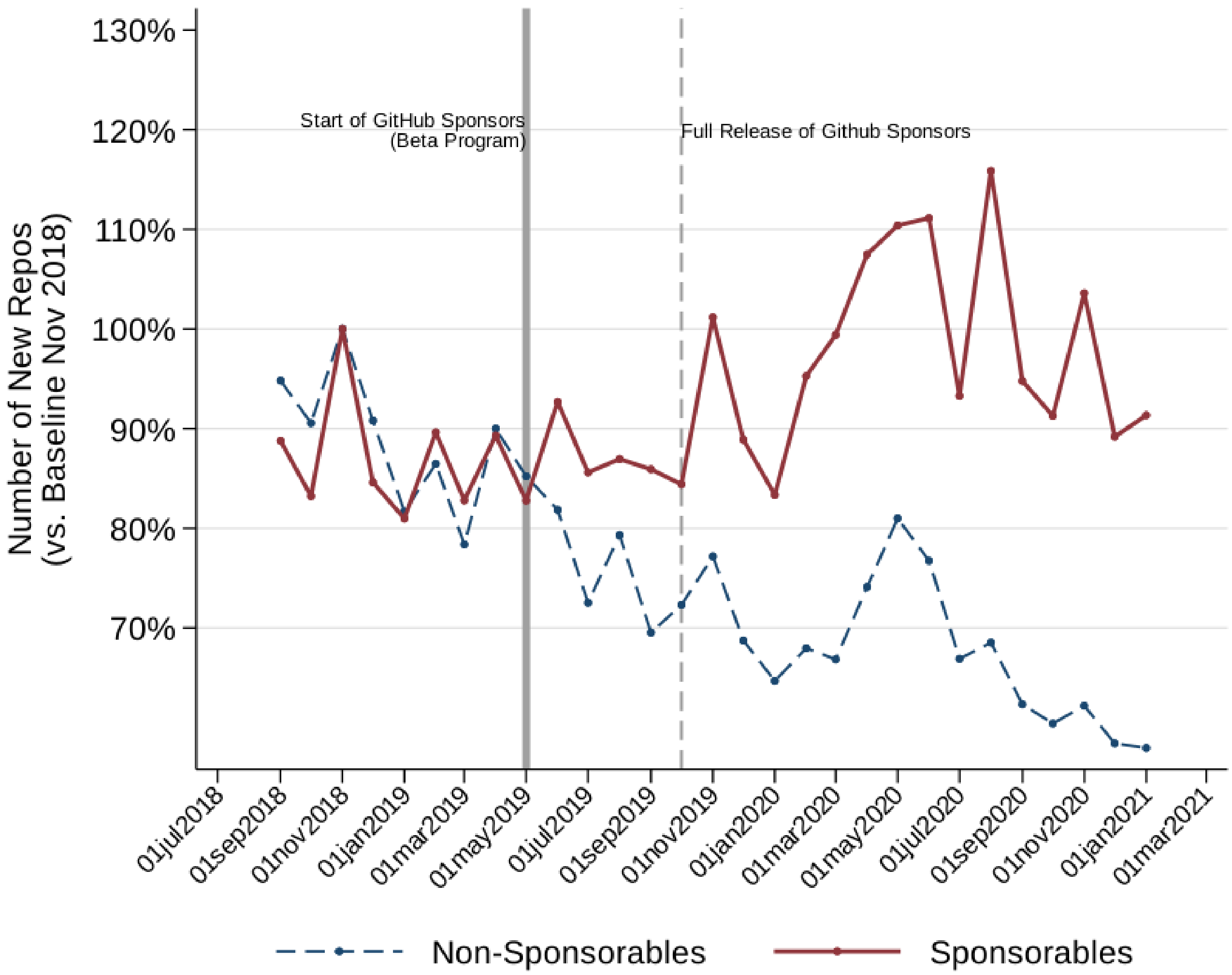
Zuegel, D. (2019), 'Announcing github sponsors: A new way to contribute to open source'.
URL: *https://github.blog/2019-05-23-announcing-github-sponsors-a-new-way-to-contribute-to-open-source/*

Figure 1: Sponsorships over time



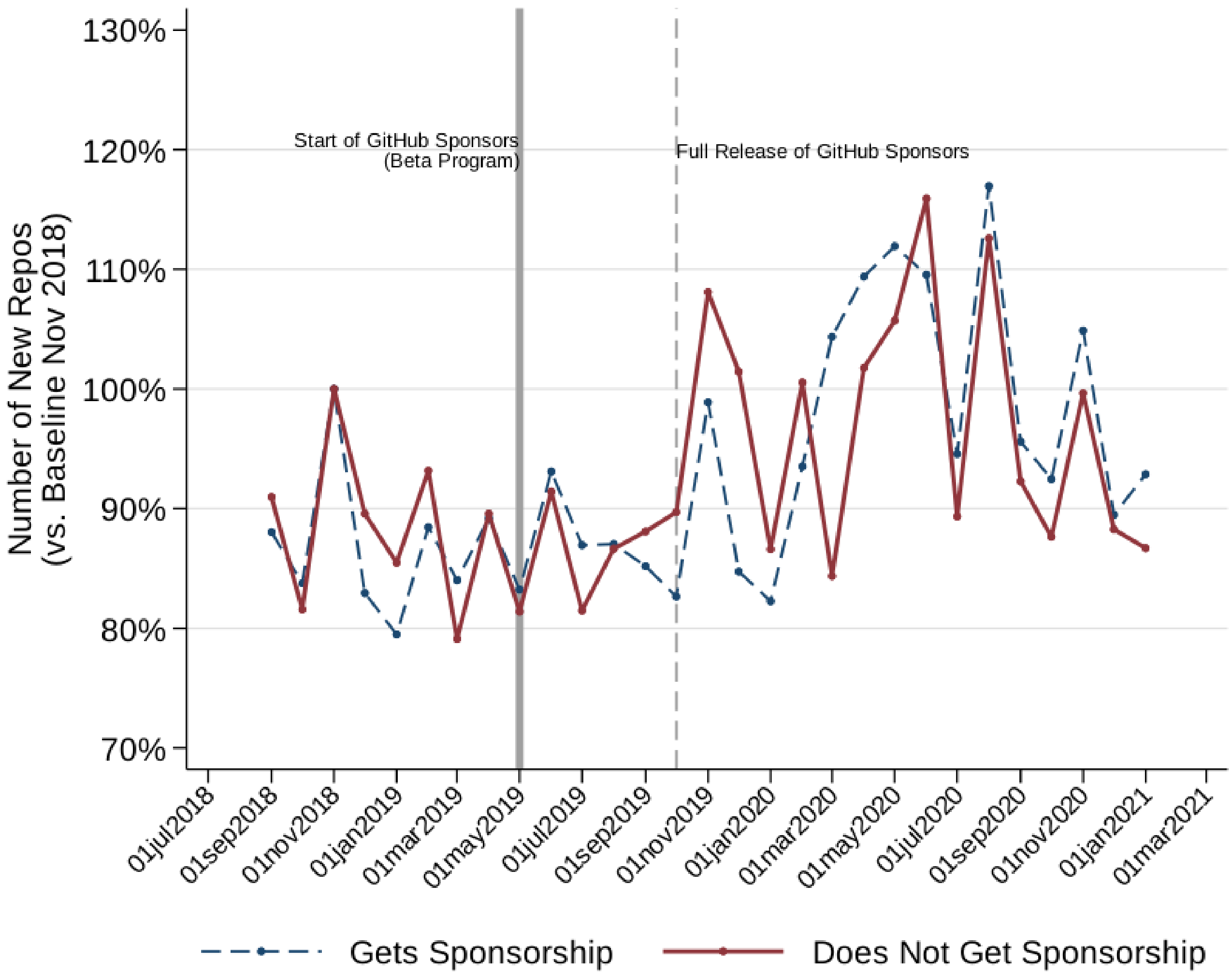
Notes: This figure depicts the evolution of number of sponsorships over time. The solid vertical line denotes the start of the program in its beta version, and the dashed vertical line the start of the full release.

Figure 2: New repositories over time distinguishing by sponsorable status



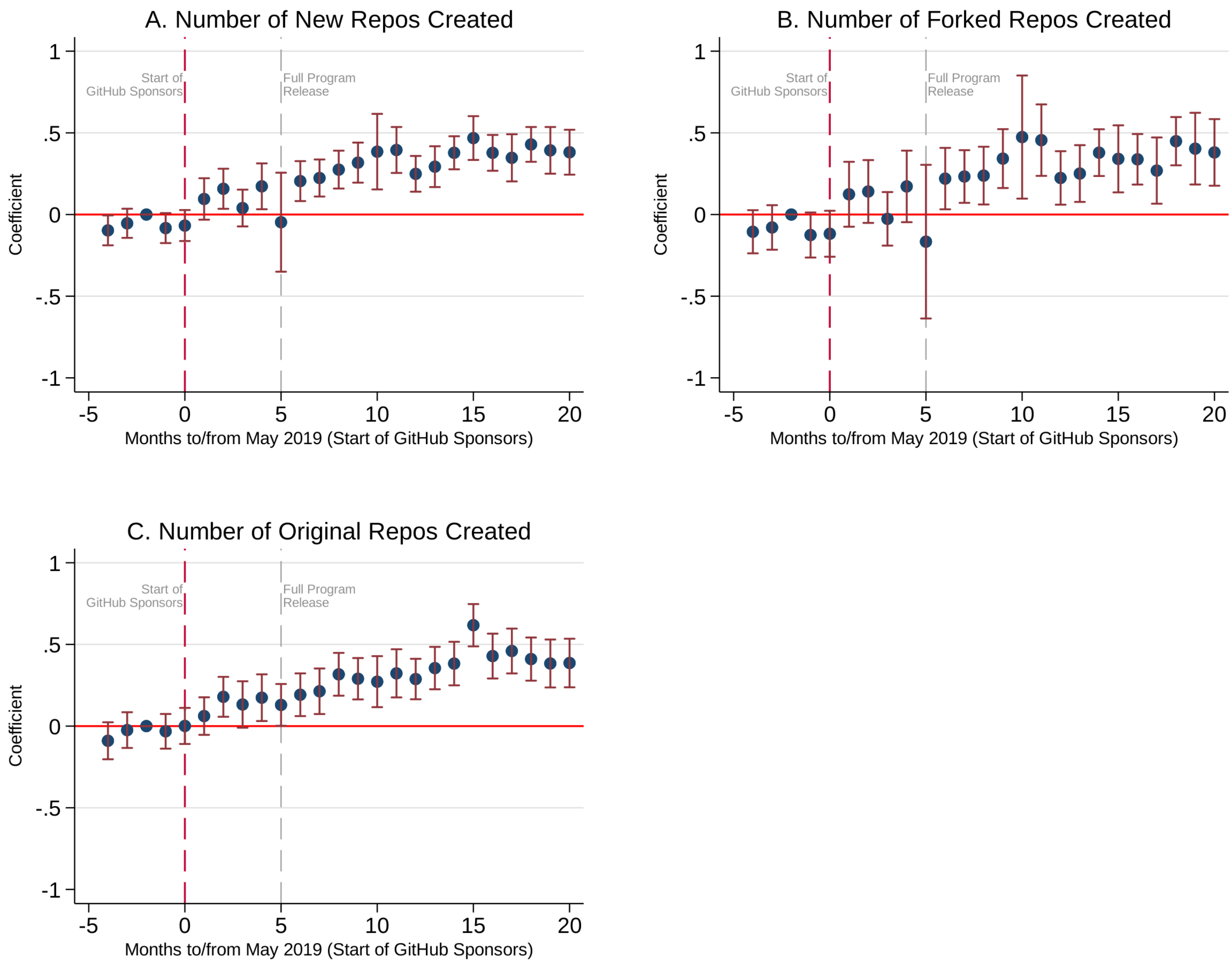
Notes: This figure depicts the evolution of the standardized number of repositories created by sponsorable and non-sponsorable users.

Figure 3: New repositories over time distinguishing between sponsored users and other sponsorables



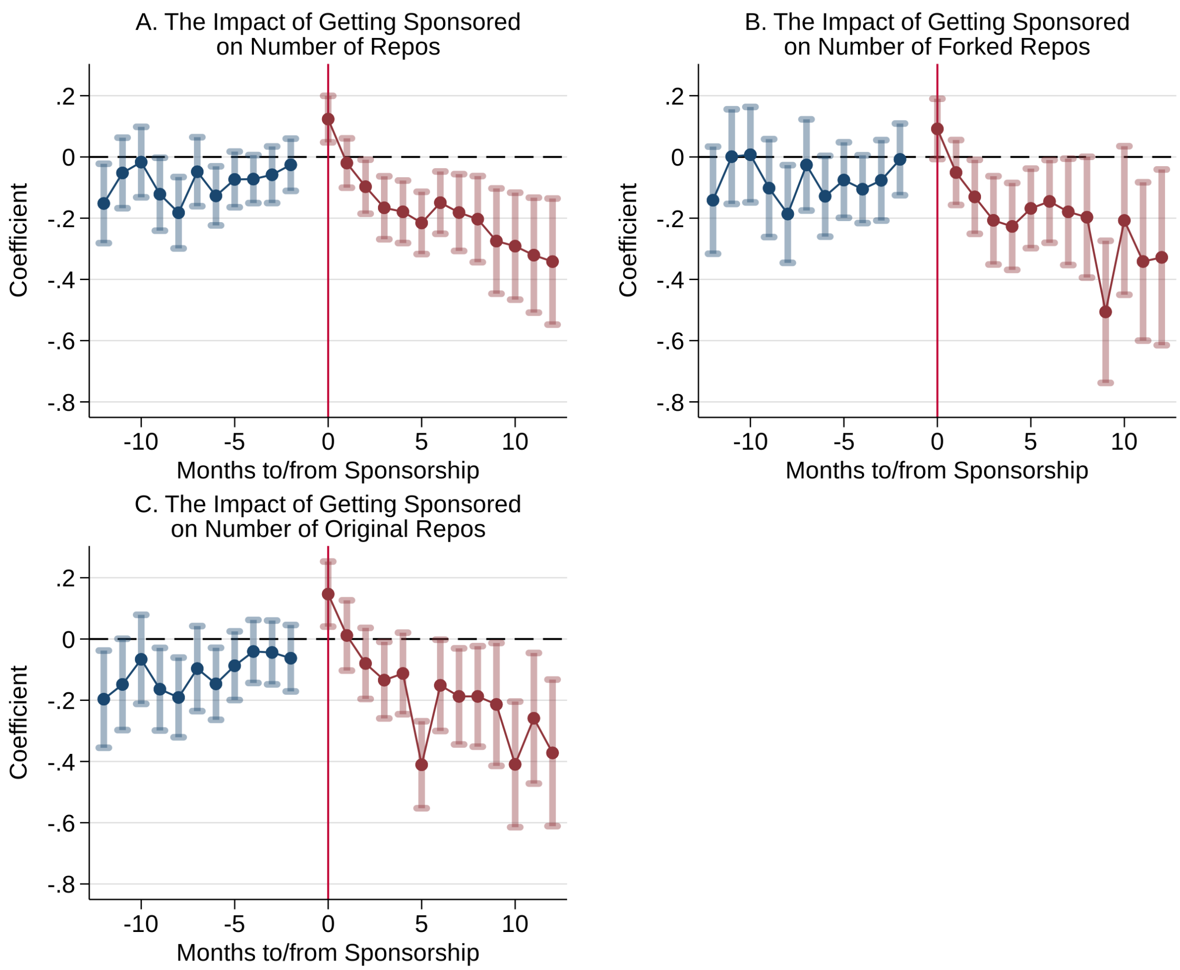
Notes: This figure depicts the evolution of the standardized number of repositories created by sponsored users and sponsorable users that were never sponsored during the sample period.

Figure 4: The impact of the launch of the Sponsorships program on sponsorables' repositories: Event study



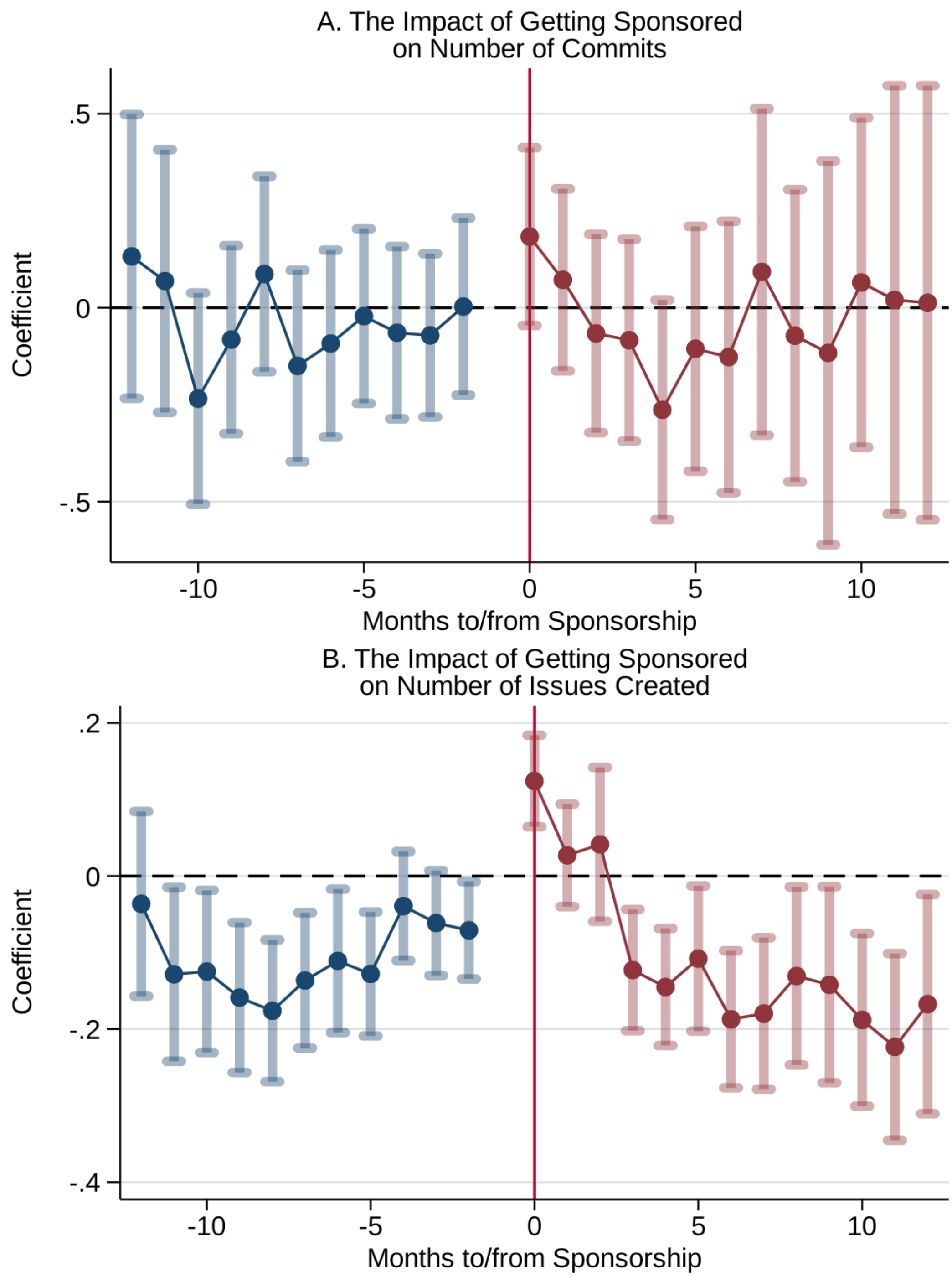
Notes: In this figure, we estimate a dynamic version of Eq. (1) that replaces $Sponsorable_i \cdot I[t > May\ 2019]$ with time indicators for each year-month to/from the launch of the Sponsors program. As an outcome, we examine the number of new repositories created by sponsorable users. In Panel A, we consider the total number of repositories, in Panel B, the number of forked repositories, while in Panel C, the number of original repositories. In all panels, each observation is weighted by the inverse probability of becoming sponsorable. User fixed-effects are included, and standard errors are clustered at the user level. Vertical bars represent 95 percent confidence intervals.

Figure 5: The Impact of receiving a first sponsorship on new repositories: Event study



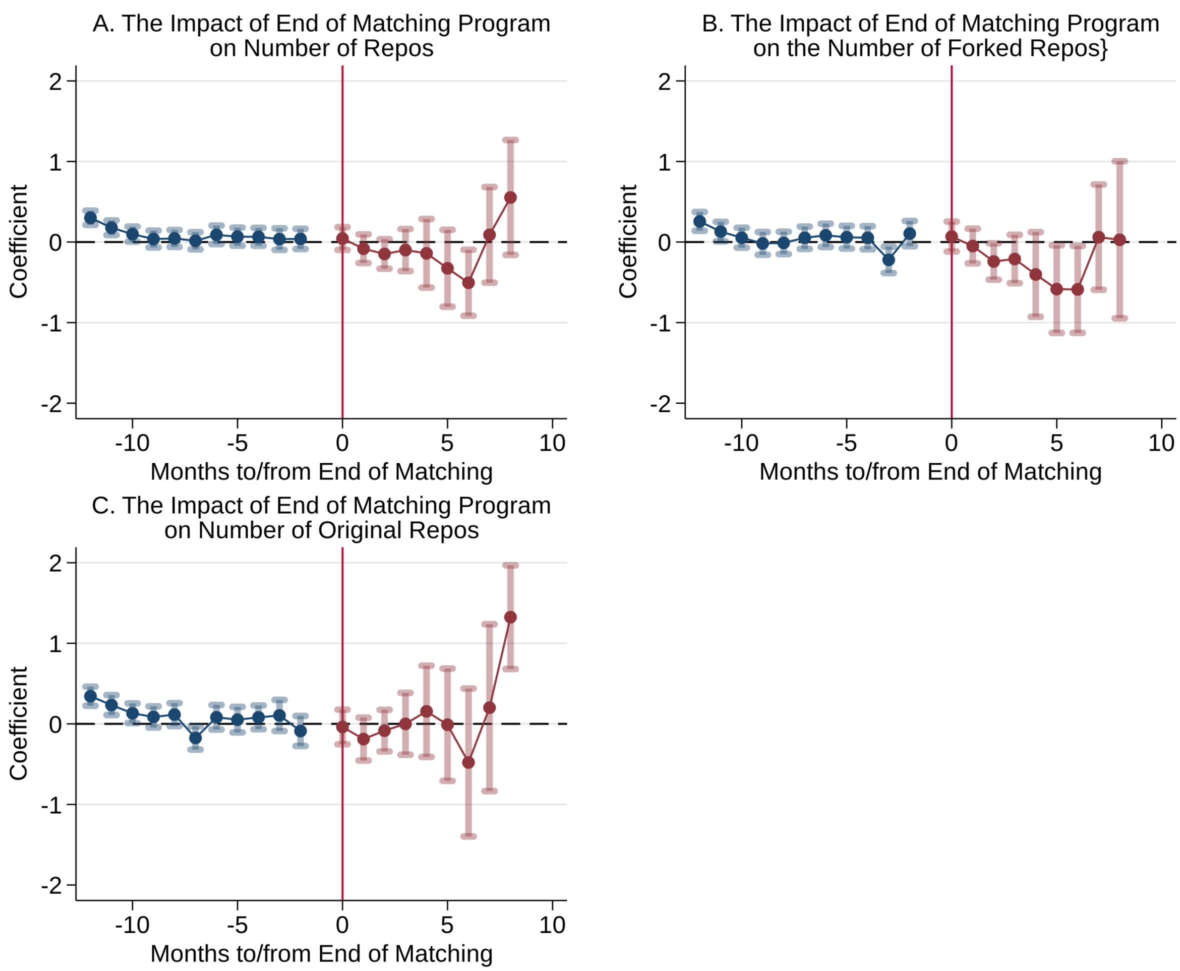
Notes: This figure reports the results of estimating a dynamic version of the Poisson model described by Eq. (2) in the main text for the number of new repositories generated. The effects associated with each month before and after obtaining a first sponsorship are computed by treatment cohort and subsequently averaged. The reference category is represented by users who were never treated during the reference period. Standard errors are clustered at the user level. Vertical bars represent 95 percent confidence intervals.

Figure 6: The impact of receiving a first sponsorship on new commits and issues: Event study



Notes: This figure reports the results of estimating a dynamic version of the Poisson model described by Eq. (2) in the main text for the number of new commits and issues created. The effects associated with each month before and after obtaining a first sponsorship are computed by treatment cohort and subsequently averaged. The reference category is represented by users who were never treated during the reference period. Standard errors are clustered at the user level. Vertical bars represent 95 percent confidence intervals.

Figure 7: The impact of the termination of GitHub financial matching



Notes: This figure reports the results of estimating a dynamic version of the Poisson model described by Eq. (2) in the main text, where the treatment is the termination of financial matching by GitHub for sponsored users. When GitHub launched the Sponsors program, it matched up to $5,000 USD of any user's sponsorship earnings for the duration of their first year in the program. Given this, we modify Eq. (2) and identify the period after one year from a user's first sponsorship as the treatment period. The effects associated with each month before and after the end of GitHub financial matching are computed by treatment cohort and subsequently averaged. The reference category is represented by users who were never treated during the reference period. Standard errors are clustered at the user level. Vertical bars represent 95 percent confidence intervals.

Table 1a: Descriptive statistics: Sponsorables vs. random non-sponsorable users

	(1) Random Users		(2) Sponsorables		(1)-(2) Pairwise t-test	
Variable	N	Mean/(SE)	N	Mean/(SE)	N	Mean difference
New Repos (all)	36197	0.334	9955	0.803	46152	-0.469***
		(0.006)		(0.025)		
New Forked Repos	36197	0.161	9955	0.447	46152	-0.286***
		(0.005)		(0.023)		
New Original Repos	36197	0.173	9955	0.356	46152	-0.183***
		(0.002)		(0.007)		
New Commits	36197	40.198	9955	156.670	46152	-116.472***
		(18.347)		(14.307)		
New Issues	36197	0.427	9955	5.083	46152	-4.656***
		(0.024)		(0.121)		

Notes: The total number of repositories ("repos"), commits, and issues are computed over the pre-sample period, which starts in September 2018 and ends in December 2018. Significance: ***=0.01, **=0.05, *=0.1.

Table 1b: Descriptive statistics: Sponsored vs. non-sponsored users

	(1) Non-sponsored		(2) Sponsored Users		(1)-(2) Pairwise t-test	
Variable	N	Mean/(SE)	N	Mean/(SE)	N	Mean difference
New Repos (all)	1180	26.597 (1.810)	1942	30.809 (0.839)	3122	-4.212**
New Forked Repos	1180	15.444 (1.373)	1942	17.872 (0.627)	3122	-2.428*
New Original Repos	1180	11.153 (0.666)	1942	12.937 (0.376)	3122	-1.785**
New Commits	1180	7442.524 (935.016)	1942	10437.049 (805.595)	3122	-2994.525**
New Issues	1180	228.228 (18.903)	1942	372.468 (11.273)	3122	-144.240***

Notes: The total number of repositories ("repos"), commits, and issues are computed over the pre-sample period, which starts in September 2018 and ends in December 2018. Significance: ***=0.01, **=0.05, *=0.1.

Table 2: The Impact of the launching of the Sponsors program on sponsorables' new repositories

	(1)	(2)	(3)	(4)	(5)	(6)
	New Repos		New Original Repos		New Forked Repos	
Is Sponsorable X After May 2019	0.431*** (0.0286)	0.238*** (0.0474)	0.464*** (0.0269)	0.298*** (0.0340)	0.409*** (0.0438)	0.193** (0.0743)
Is Sponsorable X After Beta		0.259*** (0.0540)		0.221*** (0.0376)		0.288*** (0.0847)
N	1144137	1144137	981215	981215	866984	866984

Notes: This table reports the results from estimating variants of Eq. (1) in the main text. In columns 1 and 2, the examined outcome is the number of new repositories (repos) created in month t. In columns 3 and 4, the examined outcome is the number of new original repositories created in t. In columns 5 and 6, the examined outcome is the number of new forked repositories generated in t. In columns 1, 3, and 5, *Sponsorable* is an indicator identifying sponsorable users, while $After\ May$ 2019 is a (0/1) indicator that becomes 1 from the moment the Sponsors program was launched (May 2019). In columns 2, 4, and 6, we display the results from estimating a variant of Eq. (1). Here, we add an interaction between *Sponsorable* and a (0/1) indicator, $After\ Beta$, that takes the value 1 after the end of the Sponsors' beta period (September 2019). Standard errors are clustered at the user level. Significance: ***=0.01, **=0.05, *=0.1.

Table 3: Heterogeneity by type of repositories created: Program effects

	(1) All Repos	(2) Top 5% of Forks	(3) Top 5% of Stars	(4) No Forks	(5) No Stars
Sponsorable X After May 2019	0.238***	0.204**	0.202*	0.236***	0.243***
	(0.0474)	(0.0785)	(0.0795)	(0.0301)	(0.0332)
Sponsorable X After Beta	0.259***	0.287***	0.267***	0.179***	0.182***
	(0.0540)	(0.0784)	(0.0770)	(0.0346)	(0.0379)
N	1144137	227360	229013	1115659	1093155

Notes: This table reports the results from estimating variants of Eq. (1) in the main text. in column 1, we report the baseline results for the total number of repositories generated, equivalent to the results in column 2 of Table 2. In column 2, we examine as an outcome a user's number of repositories in the 95th percentile of the distribution for the number of forks received as of December 2022. In column 3, we identify as hiqh-quality repositories those in the 95th percentile of the distribution for the number of stars obtained as of December 2022. In columns 4 and 5, the outcomes are the number of repositories with no forks and the number of repositories with no stars, respectively. *Sponsorable* is an indicator identifying sponsorable users, while $After\ May$ 2019 is a (0/1) indicator that becomes 1 from the moment the Sponsors program was launched (May 2019). $After\ Beta$, takes the value 1 after the end of the Sponsors' beta period (September 2019) and zero otherwise. Standard errors are clustered at the user level. Significance: ***=0.01, **=0.05, *=0.1.

Table 4: The impact of launching the Sponsors program on sponsorables' new commits and issues

	(1)	(2)	(3)	(4)
	New Commits		New Issues	
Sponsorable X After May 2019	0.403**	0.270	0.544***	0.310***
	(0.138)	(0.161)	(0.0454)	(0.0485)
Sponsorable X After Beta		0.173		0.311***
		(0.150)		(0.0453)
N	1109366	1109366	719490	719490

Notes: This table reports the results from estimating variants of Eq. (1) in the main text. In columns 1 and 2, the examined outcome is the number of new commits created in month t. In columns 3 and 4, the examined outcome is the number of new issues created in t. In columns 1 and 3, *Sponsorable* is an indicator identifying sponsorable users, while $After\ May$ 2019 is a (0/1) indicator that becomes 1 from the moment the Sponsors program was launched (May 2019). In columns 2 and 4, we display the results from estimating a variant of Eq. (1). Here, we add an interaction between *Sponsorable* and a (0/1) indicator, $After\ Beta$, that takes the value 1 after the end of the Sponsors' beta period (September 2019). Standard errors are clustered at the user level. Significance: ***=0.01, **=0.05, *=0.1.

Table 5: The impact of obtaining a first sponsorship

	(1) All Repos	(2) Original Repos	(3) Forked Repos	(4) Commits	(5) Issues
Post Sponsorship	-0.152***	-0.178**	-0.142**	-0.0444	-0.0879**
	(0.0424)	(0.0596)	(0.0518)	(0.131)	(0.0327)
N	1015344	1015344	1015344	1015344	1015344

Notes: This table reports the results from estimating variants of Eq. (2) in the main text for the sample of sponsorable users. The reported effects correspond to the average ATT across time and treatment cohorts. In column 1, the examined outcome is the number of new repositories (repos) created in month t. In columns 3 and 4, the examined outcomes are, respectively, the number of new original repositories and the number of new forked repositories created in t. *Post Sponsorship* is a (0/1) indicator that becomes 1 after a sponsorable user is sponsored for the first time. Standard errors are clustered at the user level. Significance: ***=0.01, **=0.05, *=0.1.

Table 6: Heterogeneity by type of repositories created: Being Sponsored

	(1) All Repos	(2) Top 5% of Forks	(3) Top 5% of Stars	(4) No Forks	(5) No Stars
Post Sponsorship	-0.152***	-0.577	-0.255	-0.236***	-0.0656
	(0.0424)	(4.259)	(0.170)	(0.0449)	(0.0530)
N	1015344	1015344	1015344	1015344	1015344

Notes: This table reports the results from estimating variants of Eq. (2) in the main text for the sample of sponsorable users. The reported effects correspond to the average ATT across time and treatment cohorts. In column 1, the examined outcome is the number of new repositories (repos) created in month t. In column 2, we examine as an outcome a user's number of repositories in the 95th percentile of the distribution for the number of forks received as of December 2022. In column 3, we identify as hiqh-quality repositories those in the 95th percentile of the distribution for the number of stars obtained as of December 2022. In columns 4 and 5, the outcomes are the number of repositories with no forks and the number of repositories with no stars, respectively. *Post Sponsorship* is a (0/1) indicator that becomes 1 after a sponsorable user is sponsored for the first time. Standard errors are clustered at the user level. Significance: ***=0.01, **=0.05, *=0.1.

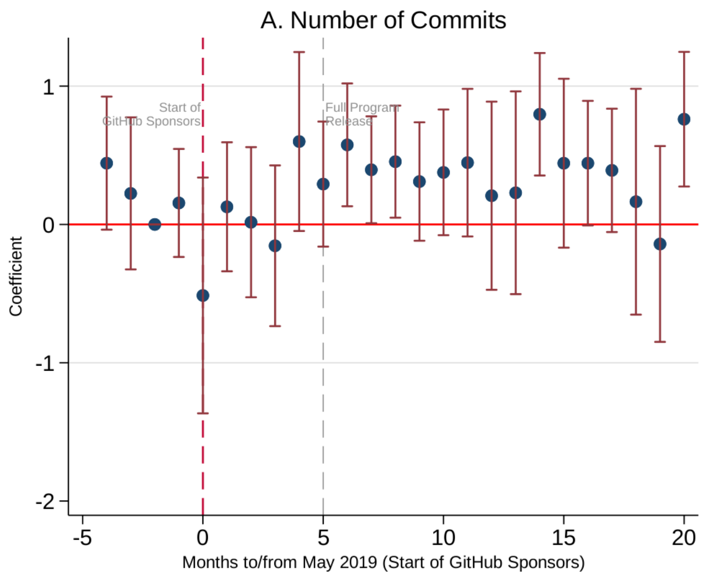
Table 7: The impact of the termination of GitHub financial matching

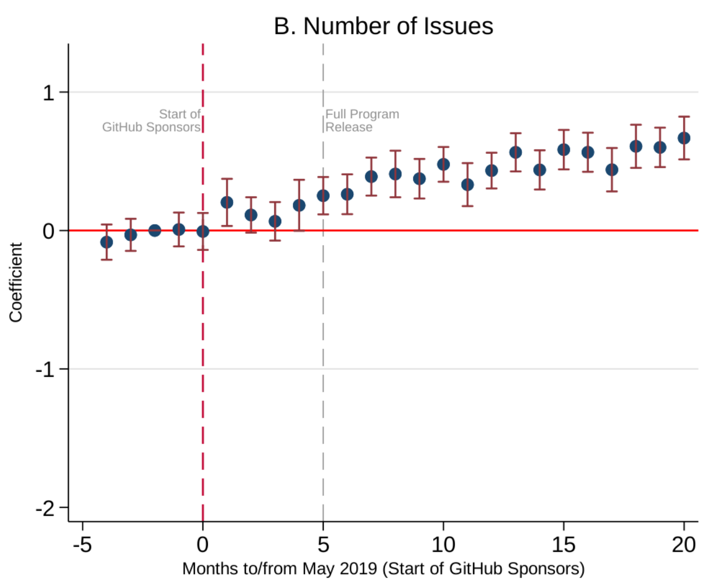
	(1) All Repos	(2) Original Repos	(3) Forked Repos	(4) Commits	(5) Issues
GitHub Matching Stops	-0.0678	-0.106	-0.0703	-0.228	0.0494
	(0.0654)	(0.0774)	(0.0997)	(0.191)	(0.0487)
N	1015344	1015344	1015344	1015344	1015344

Notes: This table reports the results from estimating a variant of Eq. (2) in the main text where the treatment is the termination of financial matching by GitHub for sponsored users. As mentioned in the paper, GitHub had matched up to $5,000 USD of any user's sponsorship earnings for the duration of their first year in the program. The reported effects correspond to the average ATT across time and treatment cohorts. In column 1, the examined outcome is the number of new repositories (repos) created in month t. In columns 3 and 4, the examined outcomes are, respectively, the number of new original repositories and the number of new forked repositories created in t. *GitHub Matching Stops* is a (0/1) indicator that is equal to one after one year from a user's first sponsorship. Standard errors are clustered at the user level. Significance: ***=0.01, **=0.05, *=0.1.

Appendix:
Incentivizing Innovation in Open Source: Evidence from the GitHub Sponsors Program

Figure A1: The impact of the launch of the Sponsors program on sponsorables' commits and issues: Event study





Notes: We estimate a dynamic version of Eq. (1) that replaces $Sponsorable_i \cdot I[t > May\ 2019]$ with time indicators for each year-month to/from the launch of the Sponsors program. As an outcome, we examine the number of new commits (Panel A) and the number of new issues (Panel B) created by sponsorable users. Each observation is weighted by the inverse probability of becoming sponsorable. User fixed-effects are included, and standard errors are clustered at the user level. Vertical bars represent 95 percent confidence intervals.