

Stack Overflow badges and user behavior

An econometric approach

Andrew Marder

Stack Overflow is a question and answer community designed for programmers. It is the largest of 130 communities in the Stack Exchange network. Created in 2008, the knowledge organized by Stack Overflow has become a valuable resource for software developers. On January 20, Spoelsky (2015) announced that Stack Exchange had raised \$40 million in venture capital funding. Stack Overflow gives users who ask questions access to expert technical help, while users who answer questions build their reputation for technical expertise and use that reputation to find better employment opportunities.

Deterding et al. (2011) define “*gamification* as the use of game design elements in non-game contexts.” Stack Overflow gamifies the process of asking and answering questions as follows. A user earns reputation points when another user votes on her posts (5 points when a question is voted up, 10 points when an answer is voted up, 15 points when an answer is accepted, and 2 points when an edit is approved). As a user earns reputation points she unlocks privileges on the site. For instance, a user must have at least 15 reputation points to vote up a question or answer.¹ Users are awarded badges for special achievements. For example, one receives the *Informed* badge by reading the tour page.²

This paper takes a first step along the path of applying econometric analysis to publicly available Stack Overflow data. Do badges motivate users to contribute to the site? Which badges are most effective? What types of user contributions are responsive to gamification? To begin answering these questions, I study how users behave around the time they are awarded badges.

METHODS

Grant and Betts (2013) present empirical evidence that three badges awarded for editing encourage recipients to make more edits in the two months preceding receipt of the badge compared to the two months after receiving the badge. This paper extends their findings by examining all types of user activity (posting questions, posting answers, and editing posts), and exploring the impact of three new badges awarded for asking questions. Table 1 describes the six badges considered in this paper.

Let y_{it} be the number of edits user i makes on day t , and t_i^* denote the day user i receives the badge of interest. Following the approach of Jacobson, LaLonde, and Sullivan (1993), I regress the number of edits user i makes on day t on a user fixed effect α_i , a set of dummy variables indicating whether the user received the badge on day $t - k$, while controlling for day of the week effects γ_j

$$\log(1 + y_{it}) = \alpha_i + \sum_{k=-29}^{30} \mathbb{1}\{t = t_i^* + k\} \delta_k + \sum_{j=1}^6 \mathbb{1}\{t \bmod 7 = j\} \gamma_j + \epsilon_{it}.$$

The model parameters are estimated using an ordinary least squares regression, and standard errors are clustered at the user level. Define $f(k)$ to be the expected number of actions taken on the k 'th day since receiving the badge

$$f(k) = E[\log(1 + y_{it}) \mid t = t_i^* + k].$$

¹A full list of privileges and the corresponding reputation points is available at <http://stackoverflow.com/help/privileges>.

²The Stack Overflow tour can be found at <http://stackoverflow.com/tour>, and all badges are listed on <http://stackoverflow.com/help/badges>.

³A question is *well-received* if it's open and has a score greater than 0. <http://meta.stackexchange.com/questions/234259/asking-days-badges>

TABLE I
BADGES OF INTEREST

Name	Description	Awarded	Introduced	Dropped
<i>Strunk & White</i>	Edited 80 posts	7,073	2008-09-15	0.00
<i>Copy Editor</i>	Edited 500 posts	1,288	2010-07-09	0.04
<i>Archaeologist</i>	Edited 100 posts that were inactive for 6 months	691	2011-08-15	0.05
<i>Curious</i>	Asked a well-received question on 5 separate days ³	138,264	2014-07-02	0.86
<i>Inquisitive</i>	Asked a well-received question on 30 separate days	14,081	2014-07-02	0.92
<i>Socratic</i>	Asked a well-received question on 100 separate days	1,240	2014-07-02	0.93
<i>Explainer</i>	Edited and answered 1 question (answer score > 0)			
<i>Refiner</i>	Edited and answered 50 questions (answer score > 0)			
<i>Illuminator</i>	Edited and answered 500 questions (answer score > 0)			

The six badges considered in this paper were introduced to the Stack Overflow site at different times. The *Strunk & White* badge was first awarded on September 15, 2008, while the three badges for asking questions were all introduced on July 2, 2014. Since the badges for asking questions were added to the site so recently many users who have been awarded these badges earned them for actions taken before the badge was introduced. For instance, 86% of users who earned the *Curious* badge were awarded the badge for actions taken before July 2, 2014. I drop these users from the analysis as they have no incentive to change their behavior to earn the *Curious* badge. The three badges awarded for editing and answering multiple questions were introduced to Stack Overflow after September 14, 2014 and are not included in the Stack Exchange Data Dump updated on September 26, 2014.

The predicted number of actions $f(\hat{k})$ is presented in Figure 1. The 95% confidence interval is depicted as a gray band around the linear prediction, standard errors were calculated using the delta method (Williams 2012).

RESULTS

The first three rows of Figure 1 illustrate how user activity changes around the time one earns a badge for editing. Each row is labeled with the name of the focal badge (*Strunk & White*, *Copy Editor*, and *Archaeologist*). There is one column for each type of user action (posting a question, posting an answer, or editing a post). The figure confirms the findings of Grant and Betts (2013). Editing increases gradually before receiving a badge for editing, with a large jump in activity on the award day. We also see that editing drops quickly after receiving the badge and gradually declines over time. It’s interesting to see how few questions were asked by the recipients of the editing badges, and to see that the rate of answering questions has a very slight increase leading up to receiving the badge and a similarly slight decrease after receiving the badge.

The results for the question-focused badges, *Curious*, *Inquisitive*, and *Socratic*, are quite different. In general, recipients of these badges are not particularly active on the site. The average level of questions, answers, and edits made all hover near zero. The uptick in questions asked on the day before receiving the badge is mechanical. Many users who earn these badges ask a question the day before they earn the badge.

By looking at a new set of badges we find that not all badges are effective at motivating user activity. The three badges for editing seem effective at changing user behavior around the time the badge is awarded. The three badges for questions do not appear effective at changing user behavior.

CONCLUSION

Stack Overflow provides a platform for job searchers to signal their ability by answering difficult technical questions publicly. Unlike Spence’s (1973) model of job market signaling, Stack Overflow enables job searchers to signal their ability in the form of a valuable public good.

When interpreting the empirical results of this paper, please consider Holland and Rubin’s motto “no causation without manipulation” (Holland 1986). There is no manipulation of the explanatory variable in this study, consequently we have not identified the causal effect of badges. To estimate the causal impact of badges on user activity we need to find a source of exogenous variation (Miller 2013).

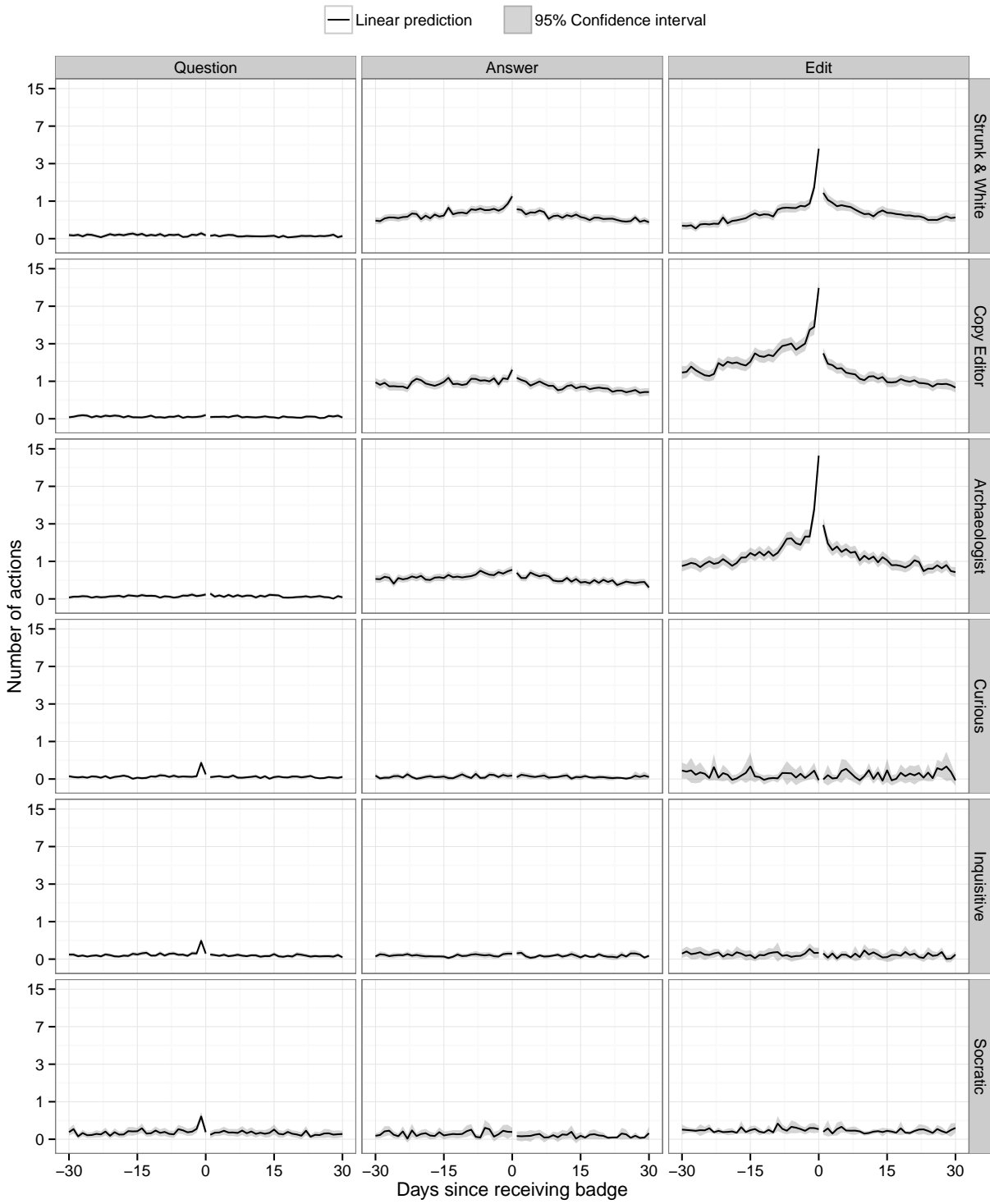


Fig. 1. User activity over time

This paper confirms the empirical observation of Grant and Betts (2013), on average users who receive a badge for editing make more edits in the 30 days prior to receiving the badge compared to the 30 days after receiving the badge. In addition, we consider how other user actions respond to editing badges. The number of questions asked each day is near zero. The slight time trends in the number of answers posted hint at a potential spillover effect, editing badges may have a secondary effect of encouraging answers. Finally, we show that users who received badges for asking questions behaved differently. In particular, we found that users do not appear motivated to change their activity levels to earn badges for asking questions.

REFERENCES

- Antin, Judd, and Elizabeth F. Churchill. 2011. "Badges in social media: A social psychological perspective." In *CHI 2011 Gamification Workshop Proceedings*. http://uxscientist.com/public/docs/uxsci_2.pdf.
- Deterding, Sebastian, Dan Dixon, Rilla Khaled, and Lennart Nacke. 2011. "From game design elements to gamefulness: Defining 'Gamification'." In *Proceedings of the 15th International Academic MindTrek Conference on Envisioning Future Media Environments - MindTrek '11*. doi:10.1145/2181037.2181040.
- Grant, Scott, and Buddy Betts. 2013. "Encouraging user behaviour with achievements: An empirical study." In *IEEE International Working Conference on Mining Software Repositories*. doi:10.1109/MSR.2013.6624007.
- Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association*. doi:10.1080/01621459.1986.10478354.
- Jacobson, Louis S., Robert J. LaLonde, and Daniel G. Sullivan. 1993. "Earnings losses of displaced workers." *The American Economic Review*. <http://www.jstor.org/stable/2117574>.
- Miller, Ashley. 2013. "Principal turnover and student achievement." *Economics of Education Review*. doi:10.1016/j.econedurev.2013.05.004.
- Spence, Michael. 1973. "Job Market Signaling." *The Quarterly Journal of Economics*. doi:10.2307/1882010.
- Spoelsky, Joel. 2015. "Andreessen Horowitz Invests in Stack Exchange." <http://blog.stackoverflow.com/2015/01/andreessen-horowitz-invests-in-stack-exchange/>.
- Stack Exchange, Inc. 2014. "Stack Exchange Data Dump." *The Internet Archive*. <https://archive.org/details/stackexchange>.
- Williams, Richard. 2012. "Using the margins command to estimate and interpret adjusted predictions and marginal effects." *Stata Journal*. <http://econpapers.repec.org/RePEc:tsj:stataj:v:12:y:2012:i:2:p:308-331>.
- Ying, Annie T. T. 2015. "Mining Challenge 2015: Comparing and combining different information sources on the Stack Overflow data set." In *The 12th Working Conference on Mining Software Repositories*.