AARHUS UNIVERSITY

Coversheet

This is the accepted manuscript (post-print version) of the article.

Contentwise, the post-print version is identical to the final published version, but there may be differences in typography and layout.

How to cite this publication

Please cite the final published version:

Bitzer, J., Geishecker, I., & Schroeder, P. J. H. (2017). Is there a wage premium for volunteer OSS engagement? - signalling, learning and noise. *Applied Economics*, 49(14), 1379-1394.
<https://doi.org/10.1080/00036846.2016.1218427>

Publication metadata

Title:	Is there a wage premium for volunteer OSS engagement? - signalling, learning and noise
Author(s):	Bitzer, J., Geishecker, I., & Schroeder, P. J. H.
Journal:	<i>Applied Economics</i>
DOI/Link:	10.1080/00036846.2016.1218427
Document version:	Accepted manuscript (post-print)

General Rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Is there a wage premium for volunteer OSS engagement? - Signaling, Learning, and Noise

Jürgen Bitzer* Ingo Geishecker[†] Philipp J.H. Schröder[‡]

October 2015

Abstract

Volunteer-based open-source production has become a significant new model for the organization of software development. Economics often pictures this phenomenon as a case of signaling: Individuals engage in the volunteer programming of open-source software (OSS) as a labor-market signal resulting in a wage premium. Yet, this explanation could so far not be empirically tested. The present paper fills this gap by estimating an upper-bound composite wage premium of voluntary OSS contributions and by separating the potential signaling effect of OSS engagement from other effects. Although some 70% of OSS contributors believe that OSS involvement benefits their careers, we find no actual labor market premium for OSS engagement. The presence of other motives such as fun of play or altruism render OSS contributions too noisy to function as a signal.

JEL: J31, J24, D01

Keywords: signaling; open-source software; voluntary work; peer production; wage formation

*Department of Economics, Carl von Ossietzky University Oldenburg, 26111 Oldenburg, Germany, e-mail: juergen.bitzer@uni-oldenburg.de

[†]Faculty of Business Administration and Economics, European University Viadrina, 15230 Frankfurt (Oder), Germany, e-mail: geishecker@europa-uni.de

[‡]Corresponding author, Philipp J.H. Schröder, Department of Economics and Business Economics, Aarhus University, Fuglesangs Allé 4, 8210 Aarhus, Denmark, Tel: +45 87164971, e-mail: psc@econ.au.dk

1 Introduction

By now open-source-based production methods, also referred to as peer production, are well established alternatives for the organization of information production and most prominently software development. The emergence of this development model, with its reliance on volunteer contributors, opens a range of fundamental questions within the field of economics and management (see Puranam et al. 2014 for an instructive account or Benkler 2011 for some implications from relaxing assumptions of pure self-interest), and is hence the subject of extensive research. See for example the works by Lerner and Tirole 2002, Mustonen 2005, Roberts et al. 2006, Shah 2006, Sauer 2007, Leppamaki and Mustonen 2009, Giuri et al. 2010, Scotchmer 2010, Fershtman and Gandal 2011 and Alexy et al. 2013, to mention only a few.

One of the most intriguing questions that OSS poses to mainstream economics and management science – and an issue of great interest to the social sciences at large – is the question of what motivates OSS programmers to donate valuable software code free of charge. A large and growing literature investigates these motives (see von Krogh et al. 2012 and Martinez-Torres and Diaz-Fernandez 2014 for recent reviews).

The majority of papers dealing with volunteer contributor motivation agree that a variety of intrinsic and extrinsic motives are at work simultaneously. Motives discussed in the literature range from intellectual curiosity to community reputation. For examples, Shah (2006) and Bitzer et al. (2007) argue that the aspect of fun and play in programming and the developers own need for software are important motives, Athey and Ellison (2014) focus on altruism and its significant role for the open source based production model, and von Engelhardt and Freytag (2013) show in a cross-country study that culture and institutional characteristics influence OSS activity. Moreover, the literature also emphasizes that extrinsic motives – like the classic economic incentive of future pay, career enhancement, or learning – explain at least part of the phenomenon (see, e.g., Hertel et al. 2003, Roberts et al. 2006, Wu et al. 2007, Ke and Zhang 2010, Alexy 2011, Hann et al. 2013).¹ One early and perhaps

¹Another branch of literature on motivation of OSS contributors and the success or

the most prominent explanation of the OSS phenomenon, initially advocated by Lerner and Tirole (2002), is of course labor-market signaling – albeit this hypothesis could not be put to proper empirical testing until now.

On first sight, labor-market signaling fits the bill, in particular, since skills of software programmers are difficult to observe directly. Likewise a programmer’s contribution and role in a project may be largely invisible in compiled machine code, i.e., closed-source software employment. Accordingly, programmers could engage in volunteer OSS programming in order to disclose their unobservable ability to employers: a situation of classic Spence-type (1973, 1974) job-market signaling. Signaling as a motive is indeed supported by a number of empirical studies that survey active contributors. These studies typically find that the desire to accelerate one’s career is repeatedly stated as a motivating factor by those individuals engaged in OSS programming (see, e.g., Hars and Ou 2002, Hertel et al. 2003, Lakhani and Wolf 2005, Wu et al. 2007, Long 2009, Bitzer and Geishecker 2010). However, such evidence cannot establish if job-market signaling is at the foundation of the open-source-based production method. Programmers stating signaling as a motive is one thing, for it to drive the OSS development process requires much more. Only when OSS engagement results in actual pecuniary rewards, it can be a signal. Put differently, OSS engagement must help employers to screen workers, i.e. resulting in a separating equilibrium. Without such a separation, other motives must be present and drawn upon in order to grasp the fundamental mechanisms of the open-source production process. Obviously, to conduct such an analysis requires ample information on programmer pay and career paths, as well as a sizable control group of non-OSS programmers. To the best of our knowledge, previous research has not featured such a controlled study. The present paper fills this gap. We are able to correctly specify the wage premium associated with voluntary OSS contributions and to fully separate the wage effects that

failure of OSS projects focusses on the relation between different forms of motivation and performance (Ke and Zhang 2009, Lee and Kim 2013), the importance of leadership style, fairness, and management in motivating OSS contributors (Oreg and Nov 2008, Dahlander and O’Mahony 2011, Li et al. 2012, Franke et al. 2013), and that the OSS development model can serve as a commitment device and facilitator for knowledge sharing (Scotchmer 2010, Fershtman and Gandal 2011, Alexy et al. 2013).

are due to signaling (i.e., from selecting into OSS) from the direct productivity effects of OSS involvement through, e.g., learning. To be able to do this, data requirements are substantial. We compile a unique database in cooperation with the most widely read German professional computer magazine “c’t Computer Magazin für Computertechnik”.² More precisely, we augment their longstanding wage survey of German IT staff with a set of questions on individual OSS activities, e.g., hours of spare time spent on OSS projects, position in OSS projects, and so on. The survey arrives at a for us relevant cross-section of 11,849 individuals and provides rich and detailed information, not only on a broad range of individual characteristics such as job functions, educational background, work experience, income, place of residence, etc., but also on the employing firm such as size, sector, and so on. The richness of information in our data set enables us for the first time to actually estimate the effect of voluntary OSS engagement conditional on a broad range of wage determinants. Furthermore, in contrast to, e.g., Orman (2008), Giuri et al. (2010) and Hann et al. (2013), our data contain information on OSS contributors and non-contributors alike.³ On this basis we are able to estimate variants of a Mincer wage equation to identify OSS wage premiums in relation to wages of the control group of non-contributors.

Obviously, an observed wage premium would be driven by both a potential signaling effect and a potential direct effect from the signal-carrying activity – for instance, learning that raises an individual’s productivity. Accordingly, simple estimates of a wage premium for OSS activity would only constitute an upper bound for the role of signaling.⁴ According to Spence’s (1973, 1974) signaling theory, individual productivity is unobservable, and only productive individuals would invest in sending the signal. From an econometric point of view, in a wage regression, the signal-carrying variable must therefore be correlated with the error term. Hence, endogeneity of the signal-carrying variable is a necessary (although not sufficient) condition for a functioning signal. Using

²The c’t has a print run of about 370,000.

³Cf. the critics of missing counterfactuals in OSS studies by Sauer (2007).

⁴In the literature on the returns to education, the related issue of signaling versus human capital effects of education has received a great deal of attention, see the survey of Weiss (1995) or Chevalier et al. (2004) for a recent empirical investigation based on UK data.

appropriate instruments and testing for the exogeneity of OSS contributions, we can thereby verify the signaling hypothesis. If signaling is indeed taking place, the coefficient of the potential signal-carrying variable must suffer from endogeneity bias. We can quantify this bias by means of IV estimation, thereby distinguishing between direct effects (e.g., through learning) and signaling effects of voluntary OSS programming activities on wages.

The central result of our investigation is that, although some 70% of the OSS contributors in the sample think that voluntary programming benefits their careers, we are unable to identify a wage premium for OSS engagement. Accordingly, unless direct wage effects of OSS are negative and overcompensate for indirect selection effects, this result is incompatible with the idea that OSS contributions function as a labor market signal. This finding is confirmed when employing instruments and testing for the exogeneity of OSS contributions: again, we find no indication of endogeneity and thus no evidence for OSS contributions functioning as a labor market signal. Only for the small group of software developers without a completed degree we are able to identify a wage premium for volunteer OSS programming, which is, however, most likely due to a direct learning effect. These results are very robust as we test a large number of different OSS indicator variables, several variations of the sample, and a multiplicity of different model specifications.

Although the evidence shows that programmers are in part attracted to OSS activities by career motives, our paper shows that such job-market signaling does not explain the persistence of the open-source development model. On the contrary, our results suggest that other motives such as play value, altruism, and peer respect are at work. Such motives play such a central role in selecting programmers into OSS activities that they render voluntary OSS engagement too noisy a signal for it to function as a labor market signal.

The following section provides a brief theoretical framework for our empirical investigation. Section 3 discusses the data and econometric method. Section 4 presents the results and discusses potential reasons why signaling may fail. Section 5 concludes.

2 Organization of Thought

To guide our empirical investigation, we briefly illustrate some central mechanisms of signaling situations. The basic setup features workers with private unverifiable information about their individual productivity, which is heterogeneous across agents. On the labor market, firms make wage offers with the aim of hiring the most productive workers. An observable activity – say obtaining a university degree, the cost of which depends on the student’s productivity – can then serve as a signal separating higher-productivity from lower-productivity individuals. However, depending on the cost and profit structures, heterogeneity distribution, and assumptions about the market structures, the system may equally well result in a pooling equilibrium, where no separation occurs. In a related theoretical paper – also cast in an OSS context – Leppamäki and Mustonen (2009) examine the interaction of job-market signaling and OSS production, when signals constitute product market externalities, a prospect not covered by our paper. They find in line with the present framework – albeit for different reasons – that situations may occur where no separation takes place.

For the current paper we mold the general principles of signaling into a simple stylized description of the software programming industry, in line with the arguments of Lerner and Tirole (2002). In doing so, it is instructive to account for key features of OSS contributions: the possibility of several simultaneously observable signal-bearing activities (e.g., education and OSS provision), private and unobservable pay-offs from such activities (e.g., user-programmer benefits, fun), and verification costs, where firms may have to expend effort in evaluating a signal (e.g., the cost of judging an individual’s programming abilities based on his or her OSS code).

Avoiding unnecessary structure, these ideas can be consolidated into the following simple setup.⁵ A population of software programmers features n individuals, i , each of whom is identified by a unique programming cost $c_i \geq 0$, i.e., reflecting his or her programming skills/productivity. These individuals

⁵The setup presented here largely reflects the consensus view about signaling laid out by Michael Spence in his Nobel Prize lecture, Spence (2002).

may engage in various signal-carrying activities, j , such as obtaining university degrees or programming OSS. While in general, the cost of displaying various signals will depend on and increase with an agent's programming costs, c_i , additional private benefits, b_i , may, however, enter and distort signals. If there exist user-programmer benefits, play values (homo ludens payoff, Bitzer et al. 2007), or gift benefits, then the private value of the verifiable activity j , net of any potential wage premium for individual i , becomes $S_{ij}(c, b)$. The various payoffs are private knowledge, while the underlying distributions and dependencies are common knowledge. For our purposes, it is not necessary to make any assumptions about these underlying distributions or dependencies, e.g., c and b may or may not be correlated. A simple labor market is given by competitive hiring from a number of identical software firms, each of which is assumed to employ $m < n$ computer programmers. Although the individual programmers' actual costs and benefits are private knowledge, employers (firms) know the distributions, and at a cost d_j can verify whether an individual displays a certain signal.⁶ The production process is such that a firm's revenue, R , depends, inter alia, on the average programming ability of its employees, σ , and we can postulate $\frac{\partial R}{\partial \sigma} < 0$. Thus hiring the average individual at wage w would imply $\sigma = \frac{1}{m} \sum_{i=1}^m c_i = \bar{c}$ and profits $\Pi = R(\sigma) - mw$. Finally, zero profits in equilibrium and the participation constraint of individuals complete the setup. Within the above structure, a number of observations can be stated. First, in any pooling equilibrium, where firms pay the pooling wage w^P we have:

Result 1. *In a **pooling equilibrium**, the pooling wage is $w^P = \frac{1}{m}R(\bar{c})$, and individuals will display a given signal if $S_{ij}(c, b) > 0$, i.e., if their private benefits exceed their respective signal-displaying costs.*

Secondly, separation may occur along a given signal. In such a situation, it is the case that:

Result 2. *In a **separating equilibrium**, where $k < n$ individuals display a signal, l , the high-wage offer is $w^H = \frac{n}{mk}R(\frac{1}{k} \sum_{i=1}^k c_i - d_l)$, while the low-wage*

⁶The cost d_j would, in the case of OSS, for example, capture the effort required to actually obtain and evaluate a programmer's OSS code. In the case of a university degree, however, d_j may be zero.

offer is $w^L = \frac{n}{m(n-k)} R(\frac{1}{n-k} \sum_{i=k}^n c_i)$, and individuals will display signal l if $S_{il}(c, b) > w^L - w^H$, i.e., the wage premium plus private benefits exceed the respective signal-displaying costs. At the same time, all other signals j will be displayed if $S_{ij}(c, b) > 0; j \neq l$.

Of course, we cannot determine here whether equilibria exist, whether separation or pooling takes place, which signal-bearing activities are observed, what size k is, or which signals are working, resulting in separation. Answers to these questions require more structure and will in general depend on the functions, R , S , the distribution of programming costs, c_i and private benefits, b_i , the size of the verification costs, d_j , and the various dependencies or independencies of these functions and distributions (see Spence 2002). However, the general results from above provide us with a series of implications, several of which can be applied to the data for empirical investigation.

Implication 1. *In a separating equilibrium, operating signals will be associated with a wage premium.*

Implication 1 contains the classic signaling hypothesis as evoked by Lerner and Tirole (2002) and leads to the straightforward empirical prediction at the center of our analysis. In one respect, however, OSS signaling in the programmer labor market differs from classic job-market signaling. In contrast to an educational signal, which may last a lifetime, the OSS contribution is a recorded observation of programming skills for a given technology and programming challenge. Thus, in view of the rapid developments in the software industry, with short knowledge life-cycles, it is unlikely for old OSS signals to command a present-day wage premium. Accordingly, we expect to find the strongest wage premium effect for contemporaneous OSS contributions.⁷

Implication 2. *With noise from private payoffs, b_i , signal-bearing activities may be observed but need not command a wage premium. Thus signals may also appear in a pooling equilibrium.*

⁷In the empirical analysis we also test the hypothesis that the wage-effect of signaling might occur with some lag.

This is because, with unobserved but sufficiently large private payoffs, even individuals with low productivity can display the signal. Thus, by Implication 1 and 2, the observation of a potentially signal-bearing activity itself does not guarantee a situation in which the signal works. On the contrary, not only may these private benefits render a separation unstable, the observed signal-bearing – but malfunctioning – activity may only be explained by referring to the private benefits. Moreover, in a situation depicted in Implication 2, the observed signal-bearing activity will not be associated with a wage premium. However, although this is generally the case for a pooling situation, separating equilibria may feature malfunctioning signals alongside the functional signal:

Implication 3. *In a separating equilibrium with the properly working signal l , an additional signal-bearing activity j may be observed. If $S_{ij}(c, b)$ and $S_{il}(c, b)$ are positively correlated, both activities may be associated with a wage premium.*

The above implications already sum up to an empirically relevant scenario where multiple signals may be observed and can be measured and compared to individual earnings. It also immediately becomes clear that this investigation strategy requires observation of individuals who do and individuals who do not display the particular signals. Any analysis based on observations of only those who do display a given signal will fail to test for the above mechanisms. Furthermore, the above illustrates that private benefits may in principle blur the situation substantially, rendering only some of all possible signals functional, or pushing the system into the pooling equilibrium. Finally, the wage premium may be inverted: for example, with dependent b and c distributions, positive mapping would disclose those displaying a signal driven by b as unproductive (high c). In line with the effects of private benefits b , the role of direct effects from a signaling activity – such as learning effects – can be considered. It is a common simplification in models of signaling and screening that the observable activity has no effect on the underlying unobservable characteristics of individuals; e.g., it is assumed that education has no effect on the human capital and hence productivity of an individual. Obviously, these simplifications, though useful from a modeling perspective, do not hold when taking the

theory to the data. This is particularly important in the extensive literature on returns to education, where signaling effects and direct human capital effects of education need to be disentangled in order to assess the true value of education; see, e.g., Weiss (1995) or Chevalier et al. (2004). The same issue occurs in the context of voluntary OSS involvement: where individuals benefit from their OSS activity in terms of training and learning (see the studies of Hars and Ou 2002, or Wu et al. 2007), this corresponds to a reduction in c_i as a result of OSS activity. Following the extension of signaling with direct productivity (human capital) effects provided in Spence (2002), we can state:

Implication 4. *In the presence of direct effects from the signal-carrying activity (learning effects), pooling equilibria may dominate some but not all previously viable separating equilibria, and investment in the signal-carrying activity may be observed even though no separation occurs.*

Next, we briefly illustrate further implications that are less suitable for empirical investigation.

Implication 5. *Ceteris paribus, more individuals will display a working signal, j , in a separating equilibrium, compared to the number of individuals displaying j in a pooling equilibrium.*

Implication 6. *An increase in the verification cost d_j makes activity j less likely to be a proper working signal.*

Verification costs will differ by the nature of the signal in question. While verifying university degrees is relatively cheap, assessing an applicant's programmed code might be relatively more cumbersome. And finally, combining Implications 2 and 6:

Implication 7. *There exists a trade-off between verification costs, d_j , and noise from $S_{i,j}$, such that a signal associated with more noise may still work if it features lower verification costs.*

Finally, and outside the scope of the above simple framework, our empirical investigation may – and most likely is – confronted with out-of-equilibrium

situations. For example, in disequilibrium, individuals may expend effort on signal-bearing activities with a view to higher wages, although no actual wage premium ultimately materializes. Whether or not such situations qualify as signaling cannot easily be resolved.

3 Data and Estimation Technique

Our data were collected in collaboration with the German professional computer magazine “*c’t Magazin für Computer Technik*”, by expanding their annual Internet-based IT wage survey for the years 2006, 2007, and 2008 to include a subsection of questions on whether the respondent contributes to OSS projects in his or her spare time. Our data are unique compared to data used in earlier empirical studies in that we assemble information on annual wages and numerous demographic and workplace-related characteristics both for individuals who do contribute as well as for those who do not contribute to OSS projects. We restrict our sample to prime-age (18-65) men and women in full-time employment. In addition to gross annual wages, we measure all additional labor income such as premiums and bonus pay. To reduce the noisiness of the data, we recode a number of wage observations that appear to correspond to monthly instead of yearly labor income, and exclude remaining observations with implausibly low gross wages (below 1,000 euros).⁸ To insure comparability across years, labor income for 2006, 2007, and 2008 is transformed into constant prices applying the national consumer price index. Annual labor income is then transformed into monthly labor income utilizing information on the number of work months per individual. Pooled over the years 2006, 2007, and 2008, our sample contains a cross-section of 11,849 individuals. In our sample, 1,952 individuals engage in voluntary OSS contributions, of whom 434 declare a role as a project leader. Most intriguing – and in line with the results of motive studies conducted among OSS contributors alone (Hars and Ou 2002, Hertel et al. 2003, Lakhani and Wolf 2005, Roberts et al. 2006, Wu et al. 2007, and Hann et al. 2013) – 1,368 of the active OSS

⁸We also estimate all specifications with the pure unaltered data. Our key findings do not change.

contributors in our sample state that they think their voluntary OSS activities benefit their career. Descriptive statistics by OSS contributor status are reported in Table 1.

To identify a potential OSS wage premium, we estimate variants of the following Mincer wage equation (see Mincer 1974):

$$\ln W_i = \alpha + \beta OSS_i + \theta DEMOG_i + \delta WORK_i + \epsilon_i \quad (1)$$

where *OSS* denotes OSS contributor status measured in various ways, *DEMOG* a set of demographic controls such as age, gender, educational attainment, and work experience, and *WORK* a number of workplace-related characteristics such as industry, occupational field, and firm size. We assume orthogonality of the error term ϵ with respect to our control variables except *OSS*. To accommodate heteroscedasticity, standard errors are bootstrapped. According to Implication 1 in Section 2, OSS contributions should generate a positive wage premium if they function as a signal and if a separating equilibrium is feasible. In the empirical application one has to be cautious, however, when interpreting the outcome of a simple regression model. First of all, we have an identification problem similar to the one prevalent in the returns to schooling literature as surveyed in Weiss (1995). There is the theoretical possibility that OSS contributions have a direct positive, or less likely, negative wage effect that may not be due to signaling, i.e., to a selection of unobservably more productive workers into OSS activities. One possibility for a direct positive wage effect of OSS engagement arises through learning. Thus, individuals may become more productive through their OSS activity, yielding a wage premium. However, this effect has to be distinguished from a signaling effect, which refers to a selection of already more productive individuals into OSS activities; see Implication 4. Accordingly, if OSS contributions indeed work as a signal for unobserved productivity, and if unobserved productivity differences yield significant wage differences, then by definition the error term in Equation 1 must be correlated with the OSS indicator variable, i.e., OSS must be endogenous.⁹ However, endogeneity does not necessarily stem from

⁹Only if the OSS indicator variable is perfectly collinear with unobserved heterogeneity,

unobserved heterogeneity. It can also arise from simultaneity between wages and OSS contributions. Thus, even if OSS contributions are correlated with the error term, this is not sufficient to establish the relevance of signaling. Our empirical strategy is therefore twofold. In a first step, we estimate a simple OLS model deriving a composite OSS coefficient that consists of the direct wage effect of OSS contributions as well as an endogeneity bias component due to unobserved heterogeneity or simultaneity. Subsequently, in a second step similar to, e.g., Angrist and Krueger (1991), we instrument for the signal-bearing activity, in our case voluntary OSS engagement, and quantify the endogeneity bias component. Accordingly, our empirical approach differs from that of, e.g., Altonji and Pierret (2001), Altonji (2005) and Bauer, and Haisken-DeNew (2001), who achieve identification by explicitly controlling for unobserved heterogeneity, e.g., through the use of measures of fathers education, sibling wages, or qualification test scores. However, by using IV techniques, we can nevertheless decompose any potential OSS premium into a signaling and direct productivity effect. To test the sensitivity of our results with respect to distorted income measures, we estimate two alternative models. First, we apply some top- and bottom-coding to monthly labor income at the first and 99th percentile. Second, we employ median regression, which is considerably more robust to outlying observations than the ordinary least squares estimator. In addition, we perform a number of robustness tests focusing on alternative OSS indicators, interaction terms, and sub-samples. Our main findings are reported in the following section.

this correlation should be zero. This is, however, unlikely as OSS is a binary variable and unobserved productivity arguably is continuous.

TABLE 1
Descriptive Statistics

	OSS-contribution				Mean Comparison H_0 :Means ident.
	No		Yes		
	Mean	Std-Deviation	Mean	Std-Deviation	
Monthly gross income	4234.11	(2186.64)	4313.48	(5970.51)	
Age (in years)	33.55	(6.85)	32.63	(6.56)	***
Tenure (in years)	6.34	(4.85)	6.51	(4.92)	
IT-Work Experience (in years)	8.71	(5.90)	8.57	(5.79)	
Gender	0.98	(0.14)	1.00	(0.06)	***
EDU: University	0.26	(0.44)	0.27	(0.44)	
EDU: Polytechnical	0.23	(0.42)	0.20	(0.40)	***
EDU: Advanced Vocational	0.11	(0.32)	0.11	(0.31)	
EDU: Vocational	0.25	(0.43)	0.24	(0.43)	
EDU: Uni without Degree	0.09	(0.29)	0.13	(0.34)	***
EDU: No Degree	0.06	(0.24)	0.05	(0.23)	
Supervisor	0.27	(0.44)	0.32	(0.47)	***
Urban Area	0.58	(0.49)	0.58	(0.49)	
Firm Size: 1 – 10	0.07	(0.26)	0.10	(0.30)	***
Firm Size: 11 – 100	0.27	(0.45)	0.32	(0.47)	***
Firm Size: 100 – 1000	0.28	(0.45)	0.27	(0.44)	
Firm Size: > 1000	0.37	(0.48)	0.31	(0.46)	***
Foreign Work Experience	0.17	(0.38)	0.21	(0.41)	***
Satisfaction with salary (very 1-6 not at all)	2.98	(1.16)	2.97	(1.18)	
D: Bad Work Climate	0.16	(0.37)	0.15	(0.36)	
D: Bad Expectations for Future	0.10	(0.31)	0.12	(0.33)	***
D: Desire to change job	0.13	(0.33)	0.09	(0.29)	***
Observations	9897		1952		

Remarks: *, **, *** indicate H_0 of Mean Comparison Test rejected with 10%, 5%, 1% error probability

4 Results and Discussion

Table 2 presents the results from a first step empirical specification, in which the indicator variable *OSS* captures all individuals who report contributing to OSS projects during their spare time, irrespective of how much effort they actually invest. Column 1 of Table 2 shows the results of a simple OLS model, while Columns 2 and 3 show the results of outlier-robust top/bottom-coding and absolute difference model estimations. Overall, the models have a very good fit and the coefficients are generally in line with what one would expect from standard wage regressions in the literature (see, e.g., Brown and Medoff 1989, Schmidt and Zimmerman 1991). Furthermore, the coefficients of the simple OLS and the two outlier-robust regressions are fairly similar. When focusing on the simple OLS model (Table 2, Column 1), our demographic and human-capital-related control variables age, tenure, and IT work experience affect wages in a non-linear way, with a peak at about 42 and at 20 and 25 years, respectively. Our results further indicate that in our sample, males, *ceteris paribus*, earn approximately 9 percent more than females.¹⁰ In addition, we find a considerable correlation between education and wages. Respondents with a university, polytechnical, or advanced vocational degree earn about 23, 15, and 6 percent more, respectively, than respondents without any formal degree. Interestingly, respondents with a basic vocational degree earn about 3 percent less, while individuals who have at least some university experience but no formal degree earn 5 percent more than respondents with no degree at all. We also find statistically significant positive wage effects of around 6 percent for those living in an urban area. Furthermore, with respect to workplace-related control variables, we find that respondents with supervisory responsibilities *ceteris paribus* earn about 9 percent more, while respondents who have some foreign work experience have a wage premium of about 9 percent. At the same time, wages increase with firm size. Respondents in small, medium, and large firms, *ceteris paribus*, earn 9, 16, and 26 percent more than respondents in very small firms with up to 10 employees. Naturally, since we do not control for individual unobserved heterogeneity, we cannot

¹⁰ $(\exp(0.0827) - 1) * 100 = 8.6$

establish any causality, thus these findings can be driven by positive selection as well as by job and firm characteristics. Regarding the *OSS* coefficient that is of highest interest in the context of the present study, none of the model specifications in Table 2 are suggestive of any positive wage premium for voluntary OSS contributions. All *OSS* coefficients are very small, none exceed their respective standard errors, and some even take on a negative sign. Accordingly, as the composite OSS effects most likely constitute upper bounds for any potential signaling effects, these results already call into question the relevance of labor market signaling through OSS contributions. The evidence, however, is not conclusive as it is at least conceivable that direct wage effects are indeed negative potentially overcompensating for positive signaling effects.

In a second step, we therefore instrument for the OSS indicator variable to assess the scope of the potential endogeneity bias. What is required is a set of variables that has sufficient explanatory power for OSS contributions while being orthogonal to the error term in the wage regression. Our data contain two variables that constitute ideal candidates for this task. Respondents give information on whether they are knowledgeable in the Linux and/or Macintosh operating system. Arguably, Linux and Macintosh users share a certain opposition to the market leader, the MS Windows operating system, and also, today perhaps more so in the case of Linux, rely more heavily on user community support. Such preferences and attitudes may be associated with a higher propensity to contribute to OSS projects on various platforms that challenge commercial software products. Despite these considerations, the reader may object the relevance of Linux and Macintosh operating system knowledge for predicting voluntary OSS contributions. Fortunately, the predictive power of excluded instruments is very easy to verify by a simple Wald test. As the large F-Statistic for our excluded instruments (*linux* and *mac*) reported in Table 3 indicates, we indeed find strong support that knowledge of Linux and/or Macintosh operating systems is a powerful predictor for voluntary OSS contributions. Naturally, in the present case one may be more worried about orthogonality of the excluded instruments. Accordingly, if after

conditioning out individual characteristics such as, e.g., educational attainment and IT work experience, knowledge in the Linux or Macintosh operating system has explanatory power for individual wages over and above OSS contributions, the associated overidentifying restrictions would not hold. However, in our sample of IT professionals, knowledge in one or the other operating system can be considered a prerequisite for survival in the industry, and one would expect wage premiums associated with different operating systems such as Windows, Linux, or Macintosh OS to quickly disappear due to competition among IT professionals. Although one cannot strictly test all overidentifying restrictions as at least one excluded instrument has to be assumed to be orthogonal to the error, the Hansen-J statistic reported in Table 3 indicates that under this assumption at least the other overidentifying restriction holds. Hence, we proceed by carrying out a Durbin-Wu-Hausman test to search for systematic differences between a consistent GMM model and a potentially inconsistent but efficient OLS model, that is, to quantify the endogeneity bias in the coefficients estimated in Table 2.¹¹

¹¹All GMM regressions and associated tests are carried out using Stata's *ivreg2* module. Baum et al. (2003) give a detailed description of the computations and instrument validity tests.

TABLE 2
Basic Regression

	Simple OLS	OLS with Top/Bottom Cod.	Median Reg. Model	Simple OLS	OLS with Top/Bottom Cod.	Median Reg. Model
OSS	-0.0016 [0.0075]	-0.0030 [0.0069]	-0.0060 [0.0074]	0.0002 [0.0079]	-0.0012 [0.0070]	-0.0031 [0.0071]
Age	0.0506*** [0.0050]	0.0490*** [0.0045]	0.0467*** [0.0047]	0.0742*** [0.0049]	0.0726*** [0.0044]	0.0724*** [0.0044]
Age^2	-0.0006*** [0.0001]	-0.0006*** [0.0001]	-0.0005*** [0.0001]	-0.0008*** [0.0001]	-0.0008*** [0.0001]	-0.0008*** [0.0001]
Tenure	0.0040 [0.0029]	0.0037 [0.0030]	0.0026 [0.0029]	0.0029 [0.0020]	0.0026 [0.0020]	0.0010 [0.0016]
$Tenure^2$	-0.0001 [0.0001]	-0.0001 [0.0001]	-0.0000 [0.0001]	-0.0000 [0.0001]	-0.0000 [0.0001]	-0.0000 [0.0001]
IT Work Experience	0.0347*** [0.0025]	0.0345*** [0.0025]	0.0326*** [0.0026]	0.0239*** [0.0023]	0.0238*** [0.0022]	0.0214*** [0.0021]
IT Work Experience ²	-0.0007*** [0.0001]	-0.0007*** [0.0001]	-0.0006*** [0.0001]	-0.0005*** [0.0001]	-0.0005*** [0.0001]	-0.0004*** [0.0001]
Gender	0.0827*** [0.0177]	0.0815*** [0.0184]	0.0909*** [0.0223]	0.0834*** [0.0181]	0.0822*** [0.0187]	0.0815*** [0.0214]
EDU: Uni	0.2041*** [0.0136]	0.2027*** [0.0123]	0.2052*** [0.0151]			
EDU: Polytechnical	0.1422*** [0.0131]	0.1421*** [0.0126]	0.1469*** [0.0145]			
EDU: Advanced Vocational	0.0572*** [0.0143]	0.0571*** [0.0139]	0.0671*** [0.0160]			
EDU: Vocational	-0.0345*** [0.0125]	-0.0351*** [0.0123]	-0.0301** [0.0136]			
EDU: Uni No Degree	0.0519*** [0.0142]	0.0525*** [0.0133]	0.0589*** [0.0152]			
Supervisor	0.0848*** [0.0060]	0.0828*** [0.0061]	0.0728*** [0.0066]	0.0921*** [0.0062]	0.0901*** [0.0062]	0.0854*** [0.0063]
Urban Area	0.0547*** [0.0063]	0.0541*** [0.0057]	0.0533*** [0.0056]	0.0590*** [0.0066]	0.0584*** [0.0059]	0.0599*** [0.0062]
Firm Size: > 1000	0.2331*** [0.0131]	0.2311*** [0.0120]	0.2229*** [0.0125]	0.2525*** [0.0137]	0.2504*** [0.0128]	0.2508*** [0.0134]
Firm Size: 100 – 1000	0.1502*** [0.0129]	0.1501*** [0.0123]	0.1374*** [0.0123]	0.1598*** [0.0134]	0.1597*** [0.0129]	0.1576*** [0.0135]
Firm Size: 11 – 100	0.0834*** [0.0121]	0.0842*** [0.0120]	0.0816*** [0.0119]	0.0926*** [0.0127]	0.0935*** [0.0127]	0.0971*** [0.0136]
Foreign Work Experience	0.0817*** [0.0071]	0.0807*** [0.0065]	0.0693*** [0.0067]	0.0914*** [0.0073]	0.0904*** [0.0067]	0.0863*** [0.0078]
Year 2007	-0.0331*** [0.0071]	-0.0301*** [0.0062]	-0.0170*** [0.0064]	-0.0343*** [0.0071]	-0.0313*** [0.0065]	-0.0233*** [0.0067]
Year 2008	-0.0263*** [0.0069]	-0.0223*** [0.0062]	-0.0110* [0.0065]	-0.0271*** [0.0070]	-0.0231*** [0.0064]	-0.0114* [0.0065]
Constant	6.3848*** [0.0929]	6.4200*** [0.0827]	6.4543*** [0.0893]	6.0575*** [0.0907]	6.0919*** [0.0802]	6.0794*** [0.0838]
Observations	11,849	11,849	11,849	11,849	11,849	11,849
(pseudo) R^2	0.544	0.571	0.3785	0.51	0.535	0.348
Employment Field Dumm. Chi^2	607.69	685.07	41.6900	1063.36	1144.01	80.8300
Federal State Dumm. Chi^2	1020.1	1026.1	54.9700	1012.4	998.05	50.9100
Industry Dumm. Chi^2	346.4	355.5	24.8900	348.79	365.56	22.7400

Remarks: Dependent variable is $\log W$. Bootstrapped standard errors in brackets. *, **, *** indicate significance at 10%, 5%, 1% error probability. Default categories: EDU: No Degree; Firm Size: 1 – 10

Table 3
Exogeneity Test for OSS

Excluded Instruments:	Knowledge in Linux OS, Mac OS	
	With Educational Controls	Without Educational Controls
Predictive power for OSS ("first stage")	$F(2, 11786) = 217.93$ $p = 0.00$	$F(2, 11791) = 221.08$ $p = 0.00$
Kleibergen-Paap Underidentification LM Test	$Chi^2 = 417.81$ $p = 0.000$	$Chi^2 = 423.45$ $p = 0.000$
Hansen J statistic for orthogonality of instruments	$Chi^2 = 0.227$ $p = 0.6340$	$Chi^2 = 0.743$ $p = 0.3887$
Durbin-Wu-Hausman Exogeneity Test	$Chi^2 = 0.099$ $p = 0.7536$	$Chi^2 = 2.081$ $p = 0.1492$

Remarks: Based on GMM model to accommodate heteroskedasticity.

As the Durbin-Wu-Hausman test statistic in Table 3 indicates, we cannot reject the H_0 that differences in the parameters between the two models are not systematic. Accordingly, our OSS indicator variable can indeed be considered exogenous, i.e., the aforementioned endogeneity bias component is in fact zero. Thus, there is no evidence either for simultaneity or for unobserved heterogeneity bias. Summarizing, we find the direct wage premium of OSS contributions to be zero and have no evidence supporting the idea that voluntary OSS contributions work as a signal for unobserved productivity. However, according to Implication 3 in Section 2, several signals may be correlated and thereby associated with a positive wage premium. In our empirical analysis, this raises the issue of multicollinearity between the potential signals *OSS* and educational attainment. However, as reported in Table 1, despite some peculiarities, educational attainment is fairly similar between OSS contributors and non-contributors. Nevertheless, we re-estimate our baseline specifications without controlling for educational attainment to identify any potential collinearity issues. As is apparent in Columns 4-6 of Table 2, the coefficients on the *OSS* indicator variable remain very small and again never exceed their standard errors, while exogeneity cannot be rejected within reasonable confidence bounds (see Table 3). Hence, there is no evidence for multicollinearity between educational attainment and OSS contributions, and Implication 3 in Section 2 is not relevant.

We further search for evidence supporting the signaling hypothesis by differentiating between different OSS indicators, thereby testing for the robustness of our finding with respect to the measurement of OSS engagement. First of all, we condition the OSS indicator to take the value one only if respondents actually invest at least three (*OSS3*) or ten (*OSS10*) hours per week. Second, we generate an indicator variable exclusively for OSS project leaders (*OSSPL*). Third, we estimate a model with actual hours per week invested in OSS projects as a regressor (*OSS – Hours*, *OSS – Hours*²). Table 4 shows the composite coefficient estimates from this exercise, reflecting the direct effect of OSS engagement as well as any potential endogeneity bias. To save space we only report the OSS related coefficients in what follows but all models include the full set of control variables as in Table 2.

TABLE 4
Alternative OSS Measures

OSS3	-0.0002 [0.0096]			
OSS10		-0.0024 [0.0192]		
OSS-PL			0.0057 [0.0148]	
OSS-Hours				-0.0003 [0.0019]
OSS-Hours2				0.0000 [0.0001]
⋮	⋮	⋮	⋮	⋮
Observations	11,849	11,849	11,849	11,849
<i>adj.R</i> ²	0.571	0.571	0.571	0.571

Exogeneity Tests of Excluded Instruments:	OSS3 Knowledge in Linux OS	OSS10 OS, Knowledge in Mac OS	OSS-PL
Predictive Power ("first stage")	$F(2, 11786) = 143.21$ $p = 0.00$	$F(2, 11786) = 42.08$ $p = 0.00$	$F(2, 11786) = 42.93$ $p = 0.00$
Kleibergen-Paap	$Chi^2 = 278.37$	$Chi^2 = 83.87$	$Chi^2 = 85.37$
Underidentification LM Test	$p = 0.00$	$p = 0.00$	$p = 0.00$
Hansen -J Statistic	$Chi^2 = 0.26$	$Chi^2 = 0.27$	$Chi^2 = 0.27$
for Orthogonality	$p = 0.61$	$p = 0.61$	$p = 0.60$
Durbin-Wu-Hausman	$Chi^2 = 0.02$	$Chi^2 = 0.03$	$Chi^2 = 0.02$
Exogeneity Test	$p = 0.89$	$p = 0.86$	$p = 0.89$

Remarks: Dependent variable is $\log W$. Bootstrapped standard errors in brackets. *, **, *** indicate significance at 10%, 5%, 1% error probability. Complete set of control variables included as in Table 2

Again, all OSS-related coefficients are very small and never exceed their respective standard errors, and some even take on a negative sign. When testing for endogeneity bias, we again cannot reject the exogeneity assumption; hence, unobserved productivity is not correlated with OSS involvement.¹²

It is conceivable that high individual rank in merit-based contributors rankings (e.g. Apache's merit system) could yield significant wage premiums and thus arguably constitutes a functioning signal. We cannot examine this prospect directly, since our data operates across all OSS projects. In contrast, status based systems are shown to play a central role – but are confined to –

¹²We do not attempt to instrument the continuous variables $OSS-Hours$, $OSS-Hours^2$ with our binary instrumental variables.

within-organizational processes (Bianchi et al. 2012). Accordingly, a merit-based ranking across several OSS projects does not exist. Still, the role as project leader or the number of hours invested in a given OSS project will be closely correlated with the value and importance of an individual's OSS contributions. Our data shows that these two potential signals – based on rank – appear to be either not credible or difficult to observe and thus do not generate positive wage effects.

In a next step, we test whether the effects of OSS contributions are heterogeneous across the sample and therefore difficult to identify at the mean. To assess this possibility in more detail, we estimate several specifications interacting our OSS indicator variables with education, age, and firm size. However as reported in Table 5, none of the interacted estimations generated a positive, statistically significant coefficient on the OSS indicator variables.

TABLE 5
Interactions

	Basic OSS			OSS Project Leader		
OSS	0.2033			0.5927		
	[0.2273]			[0.4208]		
$OSS \times Age$	-0.0088			-0.0312		
	[0.0134]			[0.0252]		
$OSS \times Age^2$	0.0001			0.0004		
	[0.0002]			[0.0004]		
$OSS \times EDU : Uni$		0.0044			0.0182	
		[0.0156]			[0.0198]	
$OSS \times EDU : Polytechnical$		-0.0176			-0.0484	
		[0.0159]			[0.0303]	
$OSS \times EDU : AdvancedVocational$		-0.0080			-0.0032	
		[0.0257]			[0.0544]	
$OSS \times EDU : Vocational$		0.0025			0.0245	
		[0.0163]			[0.0374]	
$OSS \times EDU : UniwithoutDegree$		-0.0120			0.0086	
		[0.0223]			[0.0472]	
$OSS \times EDU : Non$		0.0490			0.0835	
		[0.0324]			[0.1159]	
$OSS \times FirmSize : > 1000$			-0.0151			-0.0069
			[0.0146]			[0.0243]
$OSS \times FirmSize : \leq 10$			0.0299			0.0372
			[0.0359]			[0.0655]
$OSS \times FirmSize : FirmSize : 100 - 1000$			0.0048			0.0038
			[0.0132]			[0.0233]
$OSS \times FirmSize : FirmSize : FirmSize : 11 - 100$			-0.0025			0.0005
			[0.0123]			[0.0241]
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
Observations	11,849	11,849	11,849	11,849	11,849	11,849
$adj.R^2$	0.544	0.544	0.544	0.544	0.544	0.544

Remarks: Dependent variable is $\log W$. Bootstrapped standard errors in brackets. *, **, *** indicate significance at 10%, 5%, 1% error probability. Complete set of control variables included as in Table 2

As a further exercise, we considerably reduce sample heterogeneity by focusing on the sub-group of software developers as opposed to the remaining IT personnel in employment fields such as network administration, consulting, and web design, yielding a reduced sample size of 4,247 observations. Arguably, the signaling effect of OSS contributions should be strongest among professional software developers. Table 6 shows the respective coefficient estimates for our basic OSS indicator. Although we find no evidence for a significant general OSS wage premium (see Column 1, Table 6), when interacting the OSS indicator with educational attainment, we find a sizable positive statistically significant wage premium of OSS contributions for respondents with no formal degree. Accordingly, in the sub-sample of software developers, voluntary OSS contributions raise wages for unskilled workers by 14 percent. Furthermore, as the test statistics reported at the bottom of Table 6 show, we cannot reject the H_0 of exogeneity of OSS within reasonable confidence bounds. Accordingly, the estimated OSS wage premium for unskilled workers can most likely be accounted for by direct productivity-increasing effects (i.e., learning) and thus is not attributable to signaling. Thus, our results lend some support to the idea that voluntary OSS engagement can make software developers without any formal degree more productive. It is, however, important to bear in mind that these findings do not signify signaling and only apply to a small sub-group of the sample: in total we observe 27 unskilled OSS contributors.

TABLE 6
Programmer and Software Developer Sub-Sample

OSS	0.0124 [0.0113]	
$OSS \times EDU : Uni$		0.0107 [0.0211]
$OSS \times EDU : Polytechnical$		0.0040 [0.0197]
$OSS \times EDU : Advanced Vocational$		0.0407 [0.0430]
$OSS \times EDU : Vocational$		-0.0167 [0.0265]
$OSS \times EDU : Uni \text{ without Degree}$		0.0224 [0.0379]
$OSS \times EDU : Non$		0.1301** [0.0582]
$\vdots$	$\vdots$	$\vdots$
Observations	4,247	4,247
R^2	0.532	0.533

Exogeneity Tests for OSS	Knowledge Linux OS, Mac OS
Excluded Instruments:	
Predictive power	$F(2, 4197) = 74.02$
(“first stage”)	$p = 0.000$
Kleibergen-Paap	$Chi^2 = 142.57$
Underidentification LM Test	$p = 0.000$
Hansen J statistic for	$Chi^2(1) = 1.265$
orthogonality of instruments	$p = 0.492$
Durbin-Wu-Hausman	$Chi^2(1) = 0.612$
Exogeneity Test	$p = 0.4339$

Remarks: Dependent variable is $\log W$. Bootstrapped standard errors in brackets. *, **, *** indicate significance at 10%, 5%, 1% error probability. Complete set of control variables included as in Table 2.

5 Robustness Analysis

One potential concern with the previous analysis is that OSS contributions only affect labor market outcomes with some delay. This would be the case if OSS contributions needed to be observed for a longer time span to function as a credible signal. Furthermore, direct productivity-increasing effects (e.g., via learning) might occur only after contributing continuously for some time. To reflect on this, we altered the 2008 questionnaire by adding a further item asking for information on how long participants had been voluntarily contributing to OSS.¹³ In our 2008 sample, the median duration t of voluntary OSS contributions is five years, while 92 percent of all OSS contributors have contributed for more than one year. Accordingly, OSS contributions typically are not a one-off activity. With the 2008 data, we repeat the estimation of the basic specification and condition our voluntary OSS indicator on OSS duration in the interval of $t \geq 2$ and $t \geq 10$ years. Again we are not able to identify any statistically significant direct or signaling effect of past OSS contributions on individual wages. For brevity we only report results for $t \geq 3$ and $t \geq 5$ in Table 7.

In a final step, we repeat the estimations for the sub-sample of programmers and software developers (see Table 8). Here, we find a statistically significant direct effect. That is, respondents that have voluntarily contributed to OSS for 3 or more years ($t \geq 3$) or 5 or more years ($t \geq 5$), *ceteris paribus*, experience average wage premia of 4 percent. However, we still find no clear-cut evidence of a signaling effect. As the Durbin-Wu-Hausman statistic in the bottom panel of Table 8 indicates, we can reject exogeneity of voluntary OSS contributions only with an error probability of 13 and 12 percent for the two aforementioned durations of OSS contribution, respectively. Compared to our previous findings with an unconditional OSS indicator (see Table 6, these error probabilities are much smaller, suggesting that at least some signaling may be at play. Still we cannot accept endogeneity of our OSS indicator, and thus signaling, within reasonable confidence bounds. Furthermore, differentiating

¹³In the questionnaire, respondents are now asked whether they voluntarily contribute to OSS, and if so, for how long they have been doing so.

the direct effect with respect to education, we are able to show that the aforementioned average direct effect of OSS – like in the previous regressions – is mainly attributable to the sub-group of unskilled OSS contributors. This sub-group is very small: we observe only 8 such respondents in the 2008 sample. Accordingly, even if signaling might play a role, for which we find no conclusive statistical support, it only affects the tiny sub-group of OSS contributors without any formal education.

TABLE 7
OSS duration

	<i>Time_{OSS} ≥ 3 years</i>	<i>Time_{OSS} ≥ 5 years</i>
OSS	0.0075 [0.0135]	0.0156 [0.0152]
⋮	⋮	⋮
Observations	4,496	4,496
adj. R^2	0.567	0.567
Employment Field Dummies Chi^2	301.39	274.49
Federal State Dummies Chi^2	455.32	441.5
Industry Dummies Chi^2	156.82	154.82
<hr/>		
Exogeneity Tests for OSS	Knowledge Linux OS, Mac OS	
Excluded Instruments:		
Predictive power for OSS (“first stage”)	$F(2, 4435) = 75$ $p = 0.000$	$F(2, 4435) = 59.11$ $p = 0.00$
Kleibergen-Paap Underidentification LM Test	$Chi^2 = 145.489$ $p = 0.000$	$Chi^2 = 115.806$ $p = 0.000$
Hansen J statistic for orthogonality of instruments	$Chi^2 = 0.229$ $p = 0.6323$	$Chi^2 = 0.235$ $p = 0.6276$
Durbin-Wu-Hausman Exogeneity Test	$Chi^2 = 0.053$ $p = 0.8176$	$Chi^2 = 0.025$ $p = 0.8737$

Remarks: Dependent variable is $\log W$. Bootstrapped standard errors in brackets. *, **, *** indicate significance at 10%, 5%, 1% error probability. Complete set of control variables included as in Table 2.

TABLE 8
OSS duration, Programmer and Software Developer Sub-Sample

	<i>Time_{OSS} ≥ 3 years</i>		<i>Time_{OSS} ≥ 5 years</i>	
OSS	0.0414**		0.0396*	
	[0.0178]		[0.0226]	
<i>OSS × EDU : Uni</i>		0.0351		0.0310
		[0.0294]		[0.0338]
<i>OSS × EDU : Polytechnical</i>		-0.0105		-0.0102
		[0.0273]		[0.0321]
<i>OSS × EDU : AdvancedVocational</i>		0.1104*		0.0574
		[0.0629]		[0.0684]
<i>OSS × EDU : Vocational</i>		0.0482		0.0812
		[0.0720]		[0.1126]
<i>OSS × EDU : UniwithoutDegree</i>		0.0492		0.0631
		[0.0645]		[0.0752]
<i>OSS × EDU : Non</i>		0.2383***		0.2663***
		[0.0706]		[0.0948]
⋮	⋮	⋮	⋮	⋮
Observations	1,603	1,603	1,603	1,603
adj. R^2	0.540	0.542	0.539	0.541

Exogeneity Tests for OSS		Knowledge Linux OS, Mac OS	
Excluded Instruments:			
Predictive power for OSS ("first stage")	$F(2, 1555) = 27.62$ $p = 0.000$	$F(2, 1555) = 21.86$ $p = 0.000$	
Kleibergen-Paap Underidentification LM Test	$Chi^2 = 53.948$ $p = 0.000$	$Chi^2 = 43.19$ $p = 0.000$	
Hansen J statistic for orthogonality of instruments	$Chi^2 = 1.208$ $p = 0.2716$	$Chi^2 = 1.417$ $p = 0.2339$	
Durbin-Wu-Hausman Exogeneity Test	$Chi^2 = 2.289$ $p = 0.1303$	$Chi^2 = 2.394$ $p = 0.1218$	

Remarks: Dependent variable is $\log W$. Bootstrapped standard errors in brackets. *, **, *** indicate significance at 10%, 5%, 1% error probability. Complete set of control variables included as in Table 2.

Summarizing our results, we do not find strong evidence for a positive wage premium of voluntary OSS contributions. The lack of a composite OSS wage premium, which reflects direct wage effects through, e.g., learning as well as indirect selection effects, already casts doubts on the existence of a functioning labor market signal related to voluntary OSS contributions. When

instrumenting for voluntary OSS contributions, thereby allowing for potentially negative direct wage effects of OSS that may overcompensate indirect positive selection effects, such doubts are confirmed. We find no evidence for a selection of unobservably more productive workers into OSS activity in any specification. Hence, our empirical analysis does not lend any support to the signaling hypothesis outlined in Implication 1 in Section 2. The situation we observe is rather that described in Implication 2 in Section 2. Accordingly, voluntary OSS engagement is merely the result of higher private benefits, b , of contributors, and thus cannot carry a signal of higher productivity. Even if there is some positive selection of higher productivity types into OSS contributions, the noise introduced by the aforementioned private benefits, b , requires firms to monitor the quality of OSS contributions. Thus, our results are also consistent with a situation in which positive selection takes place, but in which the monitoring costs, d , are too high, rendering OSS contributions unusable for signaling purposes. In our theoretical reasoning at the outset of this paper, we did not make any assumptions about the distribution of unobserved productivity. If, conditional on all demographic and educational observables, unobserved productivity were sufficiently evenly distributed, then even with low monitoring costs, d , there would be no need for firms to use OSS contributions as an additional signal. Finally, in our theoretical reasoning, we considered different equilibria. If, however, we are in a situation out of equilibrium, then a separating equilibrium (Result 2 in Section 2) with an associated positive wage premium for contemporaneous voluntary OSS contributions may still materialize in the future.

6 Conclusion

The successful emergence of volunteer-based production, for example as an alternative way to organize software development, has triggered ample research within economics and management science. The question why programmers provide voluntarily software code is at the heart of this large and growing literature. The range of discussed motives is broad and often challenges mainstream assumptions. The question of whether at least part of this phenomenon

can be assigned to classical economic incentives, i.e. job-market signaling, is hence of particular interest.

This paper tested the hypothesis that OSS engagement is job-market signaling for the first time in a controlled econometric study. Spence (1973, 1974) argues that job-market signaling can explain why individuals acquire costly signals. The central assumption that individual productivity is unobservable applies perfectly to software developers. Their skills are difficult to evaluate because finished software is usually distributed in compiled machine code, i.e., as closed-source software. Thus, the signaling hypothesis might help to explain why highly educated software contributors donate valuable software code free of charge. The present paper is able to investigate this prospect because we employ a large-scale and novel data set created by augmenting a long-term wage survey carried out among German IT specialists with OSS questions. Our data differ from those used in previous studies in the richness of information and number of observations. In addition to the broad range of individual characteristics and employer information covered, our data also contain information on contributors to OSS and the counterfactual, i.e., the non-contributors.

Our econometric approach quantifies the overall wage premium associated with voluntary OSS contribution and subsequently separates direct wage effects from indirect selection effects associated with signaling by using instrumental variables. Our general finding is that our data does not display a statistically significant upper-bound wage effect that can be associated with direct productivity and indirect selection effects of voluntary OSS contributions. Furthermore, when instrumenting for voluntary OSS contributions, we find no evidence of endogeneity, a necessary prerequisite if OSS contributions indeed function as a labor-market signal. Thus, while the previous, mostly descriptive literature highlights signaling as one among many motivations for voluntary contributions to OSS, our study finds no support that labor-market signaling as a principle mechanism can sustain the open-source production model of software development. Our empirical findings imply that if individuals know that voluntary OSS programming does not work as a signal in the desired way – that is, if it does not result in a wage premium – other rewards must

motivate OSS contributors to work for free. Potential explanations, among others, include intrinsic motives like the fun of play (homo ludens payoff) and altruism. This, of course, would run counter to our descriptive finding that around 70 percent of OSS contributors expect a positive career advancement from their contributions. What our empirical analysis reveals is that despite widespread beliefs that OSS contributions generate positive career effects, the market in fact generally does not reward voluntary OSS contributions with higher wages – neither through wage premiums associated with signaling nor through direct productivity effects. One potential explanation is, that the other coexisting motives such as play value or altruism introduce noise and distort the job-market signals created by OSS contributions to the extent that they no longer convey meaningful information (i.e. result in a pooling equilibrium). A merit-based contributor ranking, like those in the Apache OSS projects, could potentially overcome such distortions and help employers to receive a functioning signal from OSS contributions to guide hiring and wage decisions.

References

- Angrist, J. D. and A. B. Krueger. 1991. Does Compulsory School Attendance Affect Schooling and Earnings?, *Quarterly Journal of Economics*, 106, no. 4:979-1014.
- Alexy, O. 2011. A fistful of Dollars: Are Financial Rewards a Suitable Management Practice for Distributed Models of Innovation? *European Management Review*, 8 (3): 165-185.
- Alexy, O., G. George, A.J. Salter. 2013. Cui Bono? The Selective Revealing of Knowledge and Its Implications for Innovative Activity. *Academy of Management Review*, 38 (2):270-291.
- Altonji, J. G. and C. R. Pierret. 2001. Employer Learning and Statistical Discrimination. *Quarterly Journal of Economics* 116, no. 1:313-350.
- Altonji, J. G. 2005. Employer Learning, Statistical Discrimination and Occupational Attainment. *American Economic Review* 95, no. 2:112-117.
- Athey, S and G. Ellison. 2014. Dynamics of Open Source Movements. *Journal of Economics and Management Strategy* 23 (2): 294-316.
- Bauer, T. K. and J. P. Haisken-DeNew. 2001. Employer learning and the returns to schooling. *Labour Economics* 8, no. 2: 161-180.
- Baum, C. F., M. E. Schaffer and S. Stillman. 2003. Instrumental variables and GMM: Estimation and testing. *Stata Journal* 3, no. 1:1-31.
- Benkler, Y. 2011. The Unselfish Gene, *Harvard Business Review*, 89 (7-8):77-85.
- Bianchi, A.J., S.M. Kang, D. Stewart. 2012. The Organizational Selection of Status Characteristics: Status Evaluations in an Open Source Community, *Organization Science*, 23 (2): 341–354.
- Bitzer J. and I. Geishecker. 2010. Who contributes voluntarily to OSS? An investigation among German IT employees. *Research Policy*, 39, no. 1: 165-172.
- Bitzer, J., W. Schrettl and P.J.H. Schröder. 2007. Intrinsic Motivation in Open Source Software Development. *Journal of Comparative Economics*

35, no. 1:160-169.

- Brown, C. and J. Medoff. 1989. The Employer Size-Wage Effect. *Journal of Political Economy* 97, no. 5:1027-1059.
- Chevalier, A., C. Harmon, I. Walker and Y. Zhu. 2004. Does Education Raise Productivity, or Just Reflect it? *The Economic Journal* 114:499-517.
- Dahlander, L. and S. O'Mahony. 2011. Progressing to the Center: Coordinating Project Work, *Organizational Science*, 22 (4):961-979.
- von Engelhard, S. and A. Freytag. 2013. Institutions, culture and open source, *Journal of Economic Behavior and Organization*, 95: 90-110.
- Fershtman, C. and N. Gandal. 2011. Direct and indirect knowledge spillovers: the social network of open-source projects. *RAND Journal of Economics*, 42 (1): 70-91.
- Franke, N., P. Keinz and K. Klausberger. 2013. Does this Sound Like a Fair Deal? Antecedents and Consequences of Fairness Expectations in the Individuals Decision to Participate in Firm Innovation, *Organization Science*, 24 (5): 1495-1516.
- Giuri, P., M. Ploner, F. Rullani and S. Torrisi. 2010. Skills, division of labor and performance in collective inventions: Evidence from open source software. *International Journal of Industrial Organization* 28, no. 1: 54-68.
- Hann, Il-Horn, J. A. Roberts, S. A. Slaughter. 2013. All Are Not Equal: An Examination of the Economic Returns to Different Forms of Participation in Open Source Software Communities. *Information Systems Research* 24, no. 3:520-538.
- Hars, A. and S. Ou. 2002. Working for Free? Motivations for Participating in Open-Source Projects. *International Journal of Electronic Commerce* 6, no. 3:25-39.
- Hertel, G., S. Nieder and S. Herrmann. 2003. Motivation of Software Developers in Open Source Projects: An Internet-based Survey of Contributors to the Linux Kernel. *Research Policy* 32, no. 7:1159-1177.
- Ke, Weiling and Ping Zhang. 2009. Motivations in Open Source Software

- Communities: The Mediating Role of Effort Intensity and Goal Commitment. *International Journal of Electronic Commerce* 13, no. 4: 3966.
- Ke, Weiling and Ping Zhang. 2010. The Effects of Extrinsic Motivations and Satisfaction in Open Source Software Development. *Journal of the Association for Information Systems* 11, no. 12: 784-808.
- Lakhani, K. R. and R. G. Wolf. 2005. Why Hackers Do What They Do: Understanding Motivation Effort in Free/Open Source Software Projects, *Perspectives on Free and Open Source Software*, eds. Feller, J., Fitzgerald, B., Hissam, S. and K. R. Lakhani, , Cambridge: MIT Press.
- Leppamaki, M. and M. Mustonen. 2009. Skill Signaling with Product Market Externality. *Economic Journal*, 119(539):1130-1142.
- Lee, Dongryul and Byung Cho Kim. 2013. Motivations for Open Source Project Participation and Decisions of Software Developers. *Computational Economics* 41, no. 1: 3157.
- Lerner, J. and J. Tirole. 2002. Some Simple Economics of Open Source. *Journal of Industrial Economics* 50, no. 2:197-234.
- Mincer, J. 1974. *Schooling, Experience and Earnings*. New York: National Bureau of Economic Research.
- Li, Yan, Chuan-Hoo Tan and Hock-Hai Teo. 2012. Leadership characteristics and developers motivation in open source software development. *Information & Management* 49, no. 5: 257-267.
- Long, Ju. 2009. Open Source Software Development Experiences on the Students Resumes: Do They Count? - Insights from the Employers Perspectives. *Journal of Information Technology Education* 8, pp. 229-242.
- Martinez-Torres, M.R. and M.C. Dias-Fernandez. 2014. Current Issues and Research Trends on Open-Source Software Communities. *Technology Analysis & Strategic Management*, 26 (1): 55-68.
- Mustonen, Mikko. 2005. When Does a Firm Support Substitute Open Source Programming? *Journal of Economics and Management Strategy*, 14, No. 1: 121139

- Oreg, Shaul and Oded Nov. 2008. Exploring motivations for contributing to open source initiatives: The roles of contribution context and personal values. *Computers in Human Behavior* 24, pp. 2055-2073.
- Orman, W. H. 2008. Giving It Away for Free? The Nature of Job-Market Signaling by Open-Source Software Developers. *The B.E. Journal of Economic Analysis & Policy* 8, no. 1:Article 12.
- Puranam, P. O. Alexy and M. Reitzing. 2014. What's "new" about new forms of organizing? *Academy of Management Review*, 39 (2): 162-180.
- Roberts, J. A. , Il-Horn Hann, and Sandra A. Slaughter. 2006. Understanding the Motivations, Participation, and Performance of Open Source Software Developers: A Longitudinal Study of the Apache Projects. *Management Science* 52, no. 7: 984-999.
- Sauer, R. M. 2007. Why develop open-source software? The role of non-pecuniary benefits, monetary rewards, and open-source licence type. *Oxford Review of Economic Policy* 23, no. 4:605-619.
- Schmidt, C. M. and K. F. Zimmermann. 1991. Work Characteristics, Firm Size and Wages. *Review of Economics and Statistics* 73, no. 4:705-710.
- Scotchmer, Suzanne. 2010. Openness, Open Source, and the Veil of Ignorance. *American Economic Review*, P&P 100: 165-171.
- Shah, Sonali K. 2006. Motivation, Governance, and the Viability of Hybrid Forms in Open Source Software Development. *Management Science* 52, no. 7: 1000-1014.
- Spence, A. M. 1973. Job Market Signaling. *Quarterly Journal of Economics* 87, no. 3:355-374.
- Spence, A. M. 1974. *Market Signaling: Informational Transfer in Hiring and Related Screening Processes*. Cambridge: Harvard University Press.
- Spence, A. M. 2002. Signaling in Retrospect and the Informational Structure of Markets. *American Economic Review* 92, no. 3:434-459.
- von Krogh, Georg, Stefan Haefliger, Sebastian Spaeth, and Martin W. Wallin. 2012. Carrots and Rainbows: Motivation and Social Practice in Open Source Software Development, in: *MIS Quarterly* 36, no. 2: 649-676.

- Weiss, A. 1995. Human Capital vs. Signalling Explanations of Wages. *Journal of Economic Perspectives* 9, no.4:133-154.
- Wu, C.-G., J. H. Gerlach, and C. E. Young. 2007. An empirical analysis of open source software developers' motivations and continuance intentions. *Information & Management* 44:253-262.