MIS Quarterly

RESEARCH NOTE

# PARTICIPATION IN OPEN KNOWLEDGE COMMUNITIES AND JOB-HOPPING: EVIDENCE FROM ENTERPRISE SOFTWARE

**Peng Huang**
Department of Decision, Operations, and Information Technologies, Robert H. Smith School of Business,
University of Maryland, College Park, MD 20742-1815 U.S.A. {huang@umd.edu}

**Zhongju Zhang**
Department of Information Systems, W. P. Carey School of Business, Arizona State University,
Tempe, AZ 85287-4606 U.S.A. {zhongju.zhang@asu.edu}

# Appendix A

## Identifying Temporary, Contract-Based IT Workers

While the focus of this work is on the relationship between participation in knowledge communities and career development, we note that potential biases could result from a practice that is fairly popular in the IT service industry: the use of temporary, or short-term, contract-based IT labor from IT outsourcing service companies. We observed in our sample that a small fraction of individuals of foreign origin (mostly from India) frequently change their jobs (usually multiple times during a year) and are possibly employees of some IT staffing/outsourcing companies that provide IT consulting and services to clients on a project basis. While there is no perfect way to identify the temporary contract workers, we removed those who have obtained educational degrees *exclusively* from institutes outside the United States (as most IT offshore contractors are educated outside the United States), who are associated with an IT outsourcing company at some point in their careers, and who have changed jobs at least three times during a one-year period.

# Appendix B

## Addressing Sample Selection Bias

Another source of potential bias in our estimation comes from sample selectivity. An individual's decision to participate in the SAP Community Network is likely to be endogenous, and therefore the assignment of treatment (the activities of participation in the SCN) is not random. If individuals self-select to participate in the SCN, and the error term in the participation model (such as unobserved innate ambition or capability) is correlated with the error term in the outcome model, it would produce a selectivity bias. To detect such estimation bias, we need a sample that consists of not only individuals who choose to be the SCN users, but also IT professionals who do not participate. Therefore, we construct a matching sample similar to the set of individuals in the main sample from LinkedIn. Specifically, for each individual in our sample, we search on LinkedIn to find a matching IT professional who was employed by the same company in the same job function, and had a similar position in 2004, the starting year of the sample period (or in the first year the individual entered the sample if there is late entry), but did not participate in the SAP Community Network. If an exact match (the combination of company and position) cannot be found, we find a matching professional who had a similar job function and position but was employed by a company of similar size, in the same industry and same geographic location. This results in an augmented sample with 1,779 individuals and 13,136 observations.

**Table B1. Results from Matching Sample Treatment Effects Model**

| Variables | (1) Job switch (2nd stage) | (2) Participation (1st stage) |
|---|---|---|
| Contribution | 0.007* | |
| | (0.004) | |
| Learning | -0.011*** | |
| | (0.004) | |
| SAP employee | -0.121*** | 0.412*** |
| | (0.030) | (0.056) |
| SAP employee * Contribution | -0.020** | |
| | (0.009) | |
| SAP employee * Learning | 0.026 | |
| | (0.016) | |
| College degree | 0.072** | -0.476*** |
| | (0.030) | (0.039) |
| Master's degree | 0.026 | -0.224*** |
| | (0.016) | (0.025) |
| Doctoral degree | 0.041 | -0.024 |
| | (0.032) | (0.106) |
| Tenure in current company | -0.013 | -0.180*** |
| | (0.010) | (0.012) |
| (Tenure in current company)$^2$ | 0.000 | 0.004*** |
| | (0.000) | (0.000) |
| Tenure in current position | 0.008 | 0.222*** |
| | (0.013) | (0.013) |
| (Tenure in current position)$^2$ | -0.000 | -0.006*** |
| | (0.000) | (0.001) |
| Management | -0.068*** | -0.110*** |
| | (0.010) | (0.026) |
| Non-IT function | -0.012 | -0.303*** |
| | (0.026) | (0.063) |
| Participation | 0.187 | |
| | (0.162) | |
| Constant | 0.094 | 0.892*** |
| | (0.159) | (0.284) |
| Year dummies | Yes | Yes |
| Industry dummies | Yes | Yes |
| Firm size dummies | Yes | Yes |
| Firm type dummies | Yes | Yes |
| No. of subjects | 1,779 | 1,779 |
| Observations | 13,136 | 13,136 |
| Pseudo R-squared | – | 0.071 |

**Notes:** Endogenous treatment-effects model reported.
***$p$ < 0.01; **$p$ < 0.05; *$p$ < 0.1.

To address the sample selection bias, we adopted the two-step treatment effects model developed by Maddala (1983), which estimates a first-stage probit model of the treatment equation and then incorporates the estimated hazard in a second-stage outcome equation. We present the results of the treatment effects model in Table B1. The estimate from the second-stage outcome model indicates that there is no systematic selectivity bias (selectivity correction $\lambda = -0.09$, $p = 0.355$). More importantly, correcting for sample selection bias, our major results still hold, and the estimated marginal effects of contribution and learning on job-hopping are smaller compared to the uncorrected models.

# Appendix C

## Additional Robustness Tests

We performed a series of additional tests to probe the robustness of our results with regard to model assumptions and alternative measures of some of the key variables. First, the inclusion of SAP employees in the sample of analyses might introduce estimation bias because they may systematically behave differently from other SCN participants, for example making substantially higher contribution than average users. We rerun the models using an alternative sample that excludes all SAP employees, and present the results in Column 1 (hazard model) and Column 2 (fixed effects linear probability model) in Table C1. We find our results still hold.

Second, one potential source of estimate bias may come from the mis-measurement of the *contribution/learning* variables, if the knowledge seeker lacks the expertise to judge the quality of the answers, or rewards too many knowledge contributors by marking their posts as helpful (since, unlike correct answers and very helpful answers, there is no limit on the number of helpful answers per thread). If such measurement errors exist, they would most likely result in an attenuation bias in the estimates. However, to assess how measurement error might influence our results, we perform a separate analysis using alternative measures of the *contribution/learning* variables. Specifically, we construct the variables by simply counting the number of questions that are resolved (questions that received either a correct answer or at least a very helpful answer), without weighting them by reward points. We present the results of the Cox proportional hazard model and the fixed effects linear probability model in Column 3 and 4 of Table C1. We show that our findings are robust to this alternative way of measuring the participation activities.

Third, although the Cox hazard model has been commonly used in the literature because of its flexibility, the estimates are not always efficient when compared with full parametric models, if the underlining baseline hazard function is known (Cameron and Trivedi 2005). We test an alternative parametric survival model specification using the popular exponential hazard function, and the results are presented in Columns 5 of Table C1. They provide further evidence that our results are robust to different model assumptions.

Finally, an increasingly popular approach of estimating longitudinal binary response data that accounts for unobserved heterogeneity is the generalized estimating equations (GEE) method (Wooldridge 2002). With a population-averaged approach, the coefficients of GEE estimate describe how the population-averaged response rather than one individual's response is conditioned on the covariates. In this way, the response for a given covariate is estimable without assumptions about the heterogeneity across individuals in the parameters. In Columns 6 of Table C1 we present results from a model of the population-averaged GEE estimate with logistic link function to test the main effects of learning/contribution on job-hopping. Again, we find that the results are consistent with those of our baseline models.

### *References*

Cameron, A. C., amd Trivedi, P. K. *Microeconometrics: Methods and Applications*, New York: Cambridge University Press.
Maddala, G. S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics*, New York: Cambridge University Press.
Wooldridge, J. M. 2002. *Econometric Analysis of Cross Section and Panel Data*, Cambridge, MA: The MIT Press.

**Table C1. Results of Additional Robustness Tests**

| Variables | (1) Exclude SAP employees, Cox hazard model | (2) Exclude SAP employees, FE model | (3) Alternative measures, Cox hazard model | (4) Alternative measures, FE model | (5) Exponential hazard | (6) GEE |
|---|---|---|---|---|---|---|
| Contribution | 0.040* | 0.013* | 0.099** | 0.023* | 0.039* | 0.048† |
| | (0.023) | (0.008) | (0.043) | (0.013) | (0.023) | (0.030) |
| Learning | -0.049** | -0.019** | -0.122** | -0.029* | -0.050* | -0.061* |
| | (0.024) | (0.008) | (0.052) | (0.016) | (0.026) | (0.033) |
| SAP employee | – | – | -0.673*** | -0.092 | -0.666*** | -0.731*** |
| | | | (0.179) | (0.057) | (0.151) | (0.224) |
| SAP employee * Contribution | – | – | -0.289** | -0.026 | -0.136 | -0.147 |
| | | | (0.136) | (0.031) | (0.083) | (0.096) |
| SAP employee * Learning | – | – | 0.325 | 0.025 | 0.187** | 0.216* |
| | | | (0.286) | (0.057) | (0.088) | (0.116) |
| College degree | 0.435*** | – | 0.455*** | – | 0.372*** | 0.471*** |
| | (0.093) | | (0.091) | | (0.096) | (0.121) |
| Master's degree | 0.118* | – | 0.086 | – | 0.058 | 0.091 |
| | (0.063) | | (0.062) | | (0.059) | (0.084) |
| Doctoral degree | 0.442* | – | 0.536** | – | 0.453* | 0.625 |
| | (0.251) | | (0.232) | | (0.257) | (0.396) |
| Tenure in current company | -0.210*** | 0.021* | -0.201*** | 0.028*** | -0.163*** | -0.122** |
| | (0.049) | (0.012) | (0.047) | (0.010) | (0.038) | (0.048) |
| (Tenure in current company)$^2$ | 0.007*** | 0.000 | 0.007*** | -0.001 | 0.005*** | 0.004** |
| | (0.002) | (0.001) | (0.002) | (0.001) | (0.001) | (0.002) |
| Tenure in current position | 0.175*** | 0.036*** | 0.169*** | 0.024** | 0.134*** | 0.173*** |
| | (0.052) | (0.011) | (0.050) | (0.010) | (0.042) | (0.051) |
| (Tenure in current position)$^2$ | -0.008*** | -0.002*** | -0.008*** | -0.001 | -0.006*** | -0.008*** |
| | (0.002) | (0.001) | (0.002) | (0.001) | (0.002) | (0.002) |
| Management | -0.485*** | -0.086*** | -0.507*** | -0.082*** | -0.373*** | -0.527*** |
| | (0.070) | (0.028) | (0.068) | (0.025) | (0.064) | (0.086) |
| Non-IT function | -0.068 | 0.003 | -0.042 | -0.002 | 0.051 | -0.068 |
| | (0.172) | (0.069) | (0.167) | (0.063) | (0.167) | (0.230) |
| Constant | n/a | -0.023 | n/a | 0.102 | -1.593*** | -1.128 |
| | | (0.240) | | (0.246) | (0.102) | (0.871) |
| Year dummies | n/a | Yes | n/a | Yes | n/a | Yes |
| No. of subjects | 870 | 870 | 904 | 904 | 904 | 904 |
| Observations | 5,967 | 5,967 | 6,470 | 6,470 | 6,470 | 6,470 |

**Notes:** Cox proportional hazard models in Columns 1 and 3. Linear probability model with fixed effects in Column 2 and 4. Parametric hazard models with exponential hazard in Columns 5. Population-averaged panel data GEE models with logistic link function in Columns 6. Robust standard errors in parentheses where applicable. All models include industry, firm size and firm type dummies.
***$p < 0.01$; **$p < 0.05$; *$p < 0.1$, †$p < 0.15$.