

Free/Open Source Software Adoption, Public Policies and Development Indicators: An International Comparison

Harald Schmidbauer¹, Mehmet Gençer², and V. Sinan Tunahoglu²

¹ Istanbul Bilgi University, Institute of Social Sciences, İnönü Cad. No:28 34387 Şişli, Istanbul, Turkey harald@bilgi.edu.tr

² Istanbul Bilgi University, Department of Computer Science, Kurtuluş Deresi Cad. No:47 34440 Dolapdere, Istanbul, Turkey mgencer,vst@cs.bilgi.edu.tr

Draft paper, intended for submission to “The Third International Conference on Open Source Systems”. Please forward all correspondence to Mehmet Gençer, mgencer@cs.bilgi.edu.tr

Abstract. Despite the growing body of research on inner workings of FOSS development, there are few studies on its relation with broader developments in society. In this study we have attempted a preliminary investigation of (1) how FOSS prevalence is related to economic and human development indicators of countries, (2) whether public policies regarding FOSS emerge in a consistent relation with these indicators in several clusters of countries constructed from the United Nation’s human development index, and (3) relation of software piracy to development indicators. Our results point to relative significance of non-economic factors in FOSS adoption, lack of consistent policies among public agencies, and irrelevance of non-economic factors on software piracy.

1 Introduction

The development of policies towards Free and Open Source Software (FOSS), whether public or private, requires a better understanding of how FOSS adoption interacts with various aspects of society. While the growing body of research on FOSS focuses on community, governance, and coordination of production activity[6, 4], or relations to wider software industry[3, 11], studies on relation of FOSS with broader developments in society and factors which affect its adoption are seldom[1, 2].

In this study we have attempted a preliminary investigation of these relations. We first address the following empirical questions: (1) how FOSS adoption is related to economic and human development indicators of countries, and (2) whether public policies regarding FOSS emerge in a consistent relation with these indicators in several clusters of countries. In addition (3) relation of

software piracy to development indicators is considered in order to highlight deficiencies in policies involving proprietary software. Building on our results and research elsewhere, we then discuss policy choices accounting for economic reality.

Corresponding methods for our investigation are as follows:

1. For an international comparison of FOSS adoption, we particularly tried to assess the variability in the occurrence of FOSS across countries on the basis of several variables related to human development, economics and transparency, and the Internet infrastructure. Adoption of FOSS in a country is operationalized as the number of hits in the country-specific Google search of the keyword “open source”. This variable will obviously be correlated with the size of a country in terms of its population.
2. To assess relation of public policies towards using FOSS with development indicators, we have attempted to relate it to a classification of countries we have developed on the basis of cluster analysis. FOSS usage in public agencies is operationalized through the server software used in several public bodies in countries.
3. Deficiencies in policies involving proprietary software is examined through analysis of piracy data for countries in relation to development indicators. For this purpose piracy rate data published by BSA was checked for its correlation to human development, economics and transparency indicators.

Our results point to (1)relative significance of non-economic factors in FOSS adoption, (2)lack of consistent policies among public agencies towards FOSS, and (3)strong dominance of economic factors and irrelevance of non-economic factors on software piracy. The finding that FOSS adoption is correlated to human development index, which is defined as “a process of enlarging people’s choices”[10], is supportive of the claims that public promotion of FOSS through information campaigns can indeed be effective[1]. Furthermore, we argue that given the economic conditions in relatively less developed countries, public intervention in software market through promoting open source is not only favorable[2], but possibly inevitable if sensible conditions in software market (such as reduced piracy) are desired. On the other hand lack of consistent public policies indicates lack of awareness among policy makers towards potential of open source software, in effect disengaging large communities from enjoying its benefits.

The following three sections lays out quantitative data, models and results for relating development indices to FOSS adoption, public policies towards FOSS and software piracy, respectively. A discussion of results in section5 attempts to interpret these results, followed by conclusions.

2 FOSS Adoption

Our investigation of the FOSS adoption is based on regression models with Google counts of “open source” as dependent variable, which we have deter-

mined in a survey in April 2006, and a selection of independent variables from the following list:

- population
- gdp per capita
- human development index (hdi)
- transparency index (tri)
- Internet usage
- relative Internet usage
- Internet hosts

In order to account for the impact of economic reality, we use gross domestic product (gdp) per capita, the transparency index [9] and Internet usage as proxies to the adoption of FOSS. We also relate FOSS adoption to human development. We choose the human development index (despite ongoing debates, see, for example [5, 7, 8]), which is published by the United Nations Development Program (UNDP)[10].

A full regression model, with all variables included as regressors, and fitted to 116 cases (countries) for which the entire data set was available, has an explanatory power (i.e., R^2) of 67%, but leaves many variables insignificant. Using stepwise procedures to fit a reduced model with fewer, and significant variables, leads to the following model, fitted to 129 cases (The numbers in parenthesis are the t values of the estimates):

$$\begin{aligned} \log(\log(\text{count.open.source})) = & -1.546 + 0.8155 \times \log(\log(\text{population})) \\ & (-2.13) \quad (3.27) \\ & + 0.0485 \times \log(\text{internet.hosts}) \\ & (4.11) \\ & + 1.3070 \times \text{hdi} \\ & (5.44) \end{aligned}$$

This model also has an R^2 of about 67%, while retaining only significant regressors. A comparison of the regressors in this model with the list of variables in the table above reveals that Internet usage, relative Internet usage, the transparency index, and gdp per capita are not significant in explaining the variability of Google counts of “open source” across countries.

It may come as a surprise that gdp per capita has so little explanatory power on the adoption of FOSS. To reassess this result, another regression model, with gdp per capita substituted for hdi, can be fitted to the data set. The explanatory power of such a model is substantially smaller than the model cited above; its R^2 is reduced to 61%. We can thus conclude that economic well-being can indeed explain a certain share of variability contained in the dependent variable, but the explanatory power is further enhanced by considering non-economic factors. This gives a clue about the importance of non-economic factors in accounting for the prevalence of FOSS.

3 Public Policies

Is there a pattern in the server software variety used by governmental organizations across countries? To investigate this question, we have looked into the server software choice (open source vs proprietary) in several prominent public bodies of countries; these are the bureau of statistics, the central bank, the finance regulator, the foreign ministry, and the postal services. In addition we have clustered countries with respect to the variables discussed earlier. Finally, we have tested the null hypothesis that cluster membership and the server software variety are independent

A cluster number of 5 was found optimal to account for differences between the countries. The five clusters (based on a k -means cluster analysis with standardized variables) can be summarized as follows (a classification tree will be included in the full paper):

- Clusters 1: Low relative Internet usage, mid-range hdi; mid-range open-source count reflecting the mid-range population of the country.
- Cluster 2: Low relative Internet usage, mid-range hdi; low open-source count reflecting the smaller size of the country as compared to Cluster 1.
- Cluster 3: Low relative Internet usage, mid-range hdi; higher open-source count reflecting the larger size of the country as compared to Cluster 1.
- Cluster 4: Highest relative Internet usage, highest hdi; very high open-source count, almost irrespective of the population size.
- Cluster 5: Very low relative Internet usage, very low hdi.

Figure 1 shows a classification tree. It illustrates how the cluster membership of a country is determined: if the condition on a fork is fulfilled, the left branch applies, and vice versa.

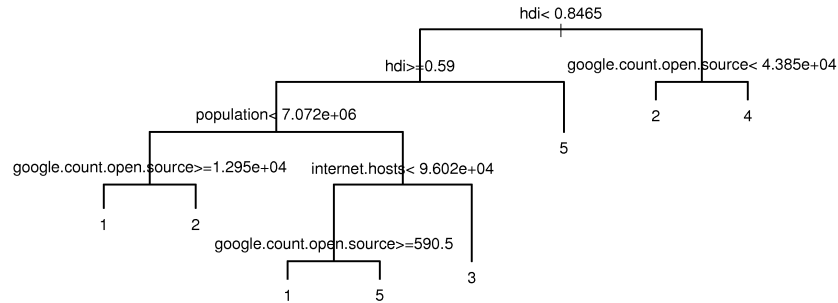


Fig. 1. Classification tree for clusters of countries

Table 1 shows the server software counts for public bodies in different clusters, with countries for which no data were available omitted. Since the main difference between Clusters 1, 2 and 3 is population size, it is justified to collapse them into a single cluster and increase the power of the χ^2 test. The p-value of the null hypothesis that ‘server software choice and cluster membership is

independent’, as shown in the last row, shows that it is rejected only in the case of the bureau of statistics

| cluster | central bank | | bureau of statistics | | finance regulator | | foreign ministry | | postal services | |
|---------|--------------|----|----------------------|----|-------------------|----|------------------|---|-----------------|----|
| | f | p | f | p | f | p | f | p | f | p |
| 1, 2, 3 | 21 | 39 | 6 | 17 | 12 | 16 | 11 | 7 | 9 | 11 |
| 4 | 9 | 12 | 6 | 18 | 13 | 11 | 16 | 6 | 14 | 10 |
| 5 | 7 | 9 | 4 | 1 | 3 | 2 | 3 | 3 | 3 | 2 |
| p-value | 0.72 | | 0.04 | | 0.63 | | 0.52 | | 0.64 | |

Table 1. Cluster membership and server software variety of organizations (f = free, p = proprietary)

A related question is: Is there evidence that the variety of server software used in one organization of a country is associated with that used in another one? We found no case in which the hypothesis of independence of server software variety in different organizations could be rejected. In this sense, we found no evidence of a consistent server software policy pattern within countries.

4 Software Piracy

The most comprehensive data for software piracy comes from Business Software Alliance(BSA), despite debates regarding its reliability. Our model attempts a linear regression with piracy as dependent variable and gdp-per-capita and transparency index(tri) as independent variables. Other models which included human development index among independent variables has failed to find any significant effect of this variable on piracy. We have found t -values of -4.385 and -4.043 for $\log(gdp.pc)$ and tri , respectively, with a R^2 value of 0.787 for the model. These results reveal dominance of economic factors on software piracy.

5 Discussion

As explained in the previous section, we found no evidence that governments pursue a server policy. This raises the question whether a top-down policy is desirable at all. Based on macro-economic models, Comino and Manenti have recently pointed that ”there may be a substantial mass of uninformed consumers, leading to market failures that may justify government intervention” [1]

Results of our empirical comparison also directs in similar ways that non-economic factors are an important determinant in harvesting possibilities FOSS provides. But more importantly it highlights that development of ”free and open source indices” of better precision can be very empowering for all parties in the game as they as their interpretation can help in policy making.

6 Conclusions

In an effort to assess the determinants of the prevalence of Free and Open Source Software (FOSS), we tried to account for the variability in Google counts of the keyword “open source” across countries worldwide using a regression model with several regressors characterizing human development, the economic situation, and the Internet infrastructure of the countries considered. It turned out that the human development index, population, and the number of Internet hosts are significant, while Internet usage is not.

A cluster analysis of the country data pointed at the interplay of human development and FOSS prevalence. We found that the bureau of statistics is the only example among several prominent organizations of a country whose server software variety (free or proprietary) is significantly related to the cluster membership of countries.

We conclude from our models that gdp per capita is less important in accounting for prevalence of FOSS, and that focusing only on increasing Internet usage will not promote the use of FOSS. It is rather information campaigns and educational incentives that may be conducive to widespread use of FOSS.

We mentioned earlier that FOSS introduced its own mode of production, in which human capital plays the most important role. The study by Streeten [8] emphasizes that “human beings are both ends in themselves and means of production”. With respect to that, we suggest that FOSS be further discussed under the aspect of its contribution to human development.

References

1. Comino S, Manenti FM (2005) Government Policies Supporting Open Source Software for the Mass Market. *Review of Industrial Organization* 26(2):217–240.
2. Dalle JM, Jullien N (2003) Libre software: turning fads into institutions? *Research Policy* 32(1):1–11.
3. Garud R, Kumaraswamy A. (1993) Changing competitive dynamics in network industries: an exploration of Sun Microsystems’ open systems strategy. *Strategic Management Journal* 14(5):351–369.
4. Hippel Ev, Krogh Gv (2003) Open source software and the private-collective innovation model: Issues for organization science. *Organization Science* 14(2):209–223.
5. Kelley AC (1991) The Human Development Index: “Handle with Care”. *Population and Development Review* 17(2):315–324.
6. Kogut B, Metiu A (2001) Open source software development and distributed innovation. *Oxford Review of Economic Policy* 17(2):248–264.
7. Srinivasan TN (1994) Human Development: A New Paradigm or Reinvention of the Wheel? *The American Economic Review* 84(2):238–243.
8. Streeten P (1994) Human Development: Means and Ends. *The American Economic Review* 84(2):232–237.
9. Transparency International (2004) Annual Report 2004 - The Coalition Against Corruption. Transparency International

10. UNDP (1990) Human Development Report 1990. Oxford University Press, New York, Oxford.
11. West J (2003) How open is open enough: melding proprietary and open source platform strategies. *Research Policy* 32(7):1259–1285.