




Which firms participate in open source software development?

A study using data from Debian

Rishab Ghosh, Kirsten Haaland, and Bronwyn H. Hall




UNITED NATIONS UNIVERSITY
UNU-MERIT



What is this paper about?

- Facts rather than hypotheses
 - i.e., curiosity driven research
- Many firms contribute to open source software
 - Which are they, and where do they come from (region, industry, size?)
- Based on Debian source code from 5 releases (1998-2005)
- Preliminary!



UNITED NATIONS UNIVERSITY
UNU-MERIT

Why do firms contribute to open source (with their own money)?*

*apologies to Nate Rosenberg

Note: Lerner and Tirole (2004) report that IBM spent >\$1B on OSS in 2001

- Enables customization to their own needs (Kuan 2002; Bessen 2002)
 - Subcase: provision of drivers for new hardware
 - Harhoff, Henkel et al (2003) – sponsored standards
- De novo entry with a new business model built on services (Dahlander 2007, *inter alia*)
- Fosfuri et al. (2003) stress pre-OSS firm assets as determinants – OSS contributors have SW tech competence and HW market presence (e.g., IBM, Sun, H-P)
- Absorptive capacity – positions the firm to take advantage of new developments in software (Rosenberg 1990, Cohen and Levinthal 1989)



Debian description (1)

- According to the website (<http://www.debian.org>):
 - “The Debian Project is an association of individuals who have made common cause to create a free operating system. This operating system that we have created is called Debian GNU/Linux, or simply Debian for short.
 - Debian GNU/Linux provides more than a pure OS: it comes with over 18733 packages, precompiled software bundled up in a nice format for easy installation on your machine.”
- Largest distribution of FLOSS software in terms of number of packages and lines of code, about 250M
 - supports a large number of hardware platforms
 - stable and mature





Debian description (2)

- Largest packages:
 - Open office
 - Kernel source 2.6.8
 - Mozilla (firefox) – web browser
 - gcc – compilers
 - Xfree86 – X window implementation
 - gimp – image manipulation
- Our data collected by researchers at U Rey Juan Carlos, Spain



Definitions

- SLOC – source lines of code
- SLOC non-duplicate – source lines of code, adjusted for code re-use (counted only once)
- Package-version – unique package name and version number - Debian version
- Package – unique package name aggregated over versions (numeric and Debian)



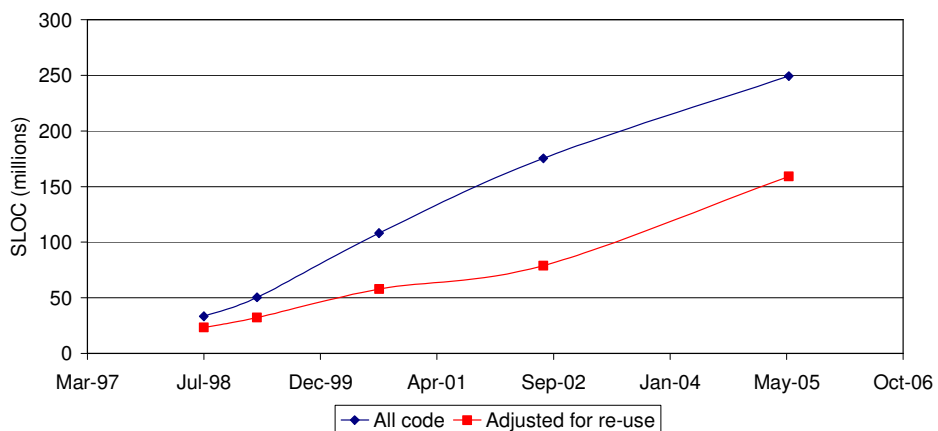


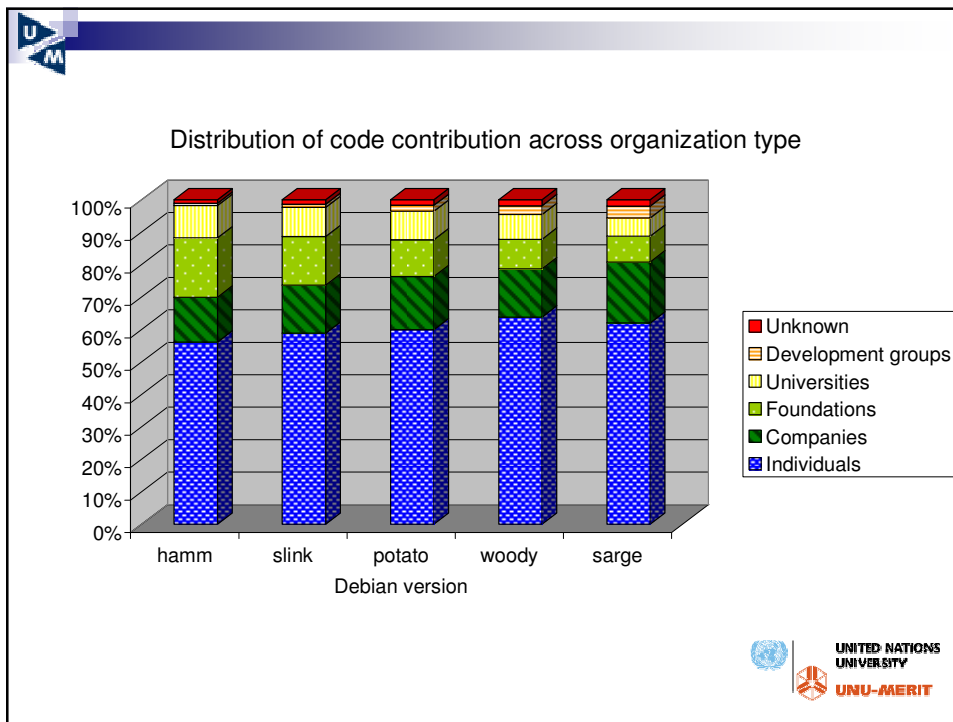
Debian versions in our data

Debian version number	Debian version name	Release date	Time between releases
2.0	Hamm	24 July 1998	
2.1	Slink	9 March 1999	200 days
2.2	Potato	14 August 2000	510 days
3.0	Woody	19 July 2002	705 days
3.1	Sarge	6 June 2005	1050 days



Contributions to Debian by version release date

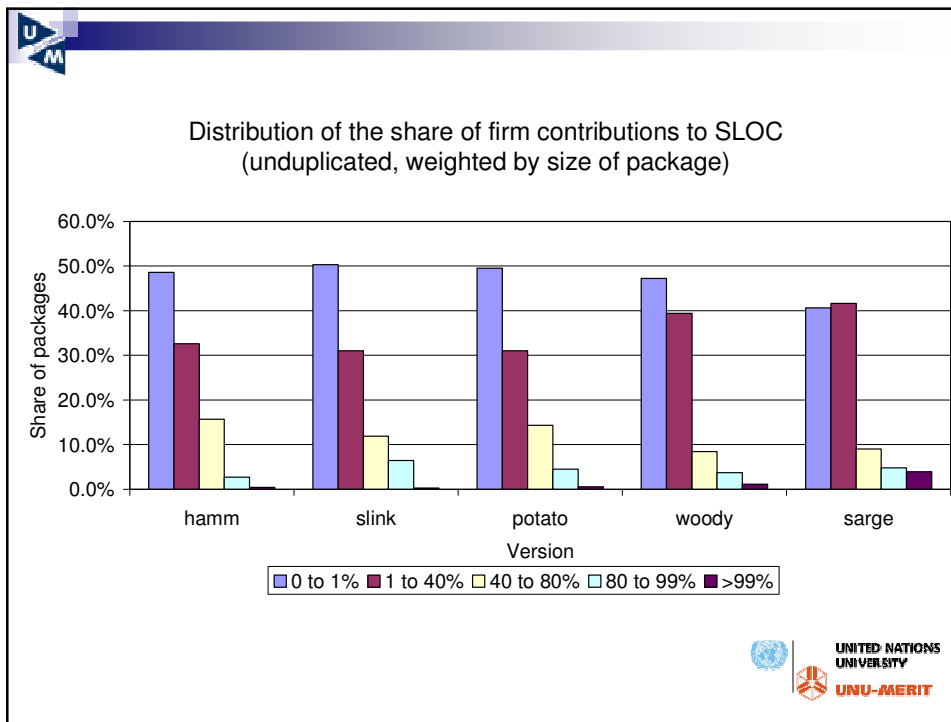




Trends in firm contributions

Version number	Version name	Number of firms	SLOC adjusted for re-use			
			Mean	Median	Min	Max
2.0	hamm	168	19,248	1714	14	804,250
2.1	slink	210	22,741	1722	5	1,129,123
2.2	potato	331	28,636	2065	5	2,651,195
3.0	woody	534	22,188	2439	6	1,258,856
3.1	sarge	903	32,927	2149	1	5,182,003
All		982	27,526	2143	1	5,182,003

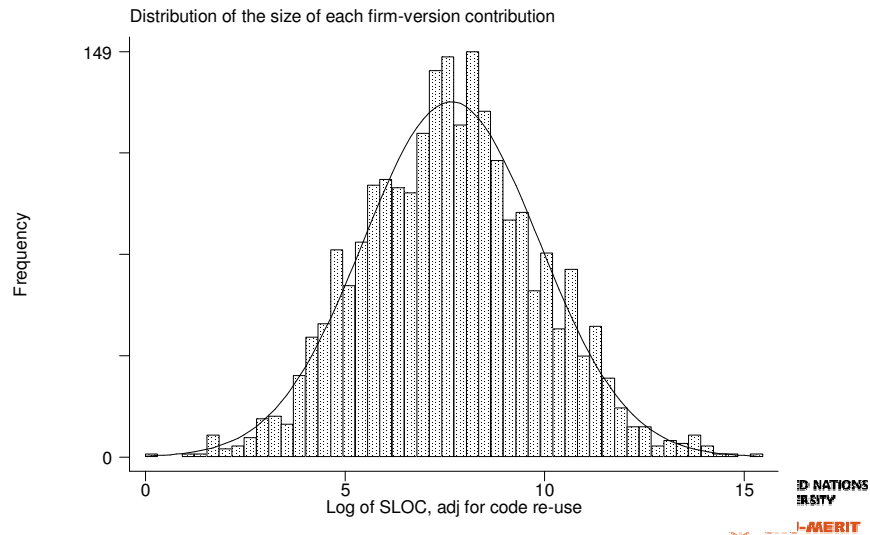
UNITED NATIONS UNIVERSITY
UNU-MERIT



- ## Extreme skewness
- Half the firms contribute to only one package and 7 firms contribute to more than 100:
 - Aladdin Enterprises, Hewlett-Packard (incl. Compaq and Dec), IBM, Red Hat, RSA Data Security, Silicon Graphics, Sun Microsystems
 - 60% of packages have contributions from one firm and 5 have contributions from more than 50:
 - kernel-image-hppa, kernel-image-ia64, kernel-source, linux-kernel-headers, xfree86
- UNITED NATIONS UNIVERSITY
UNU-MERIT



Nevertheless, distribution looks log normal



Next few slides

- Characteristics of the firms that contribute, by
 - Employment size (<10, 11-20, 21-99, 100-499, 500+)
 - Sector (various software/hardware sectors, other services and manufacturing)
 - Region (North America, Europe, Asia, Oceania, Africa & Mideast, Latin America)

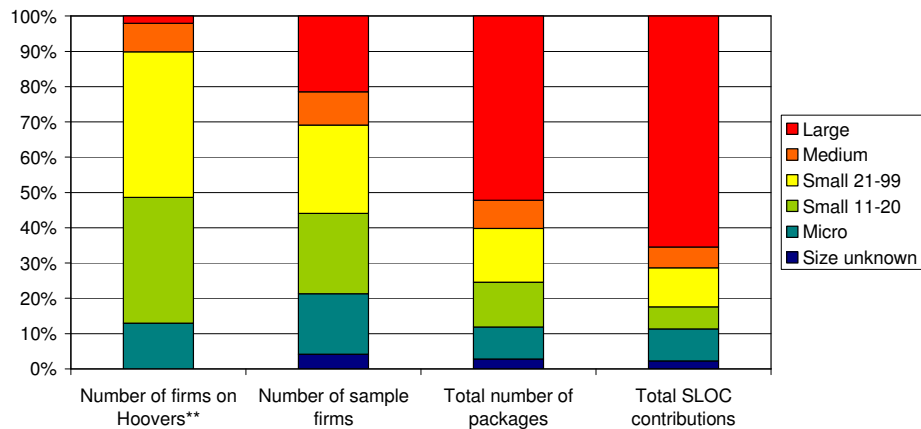


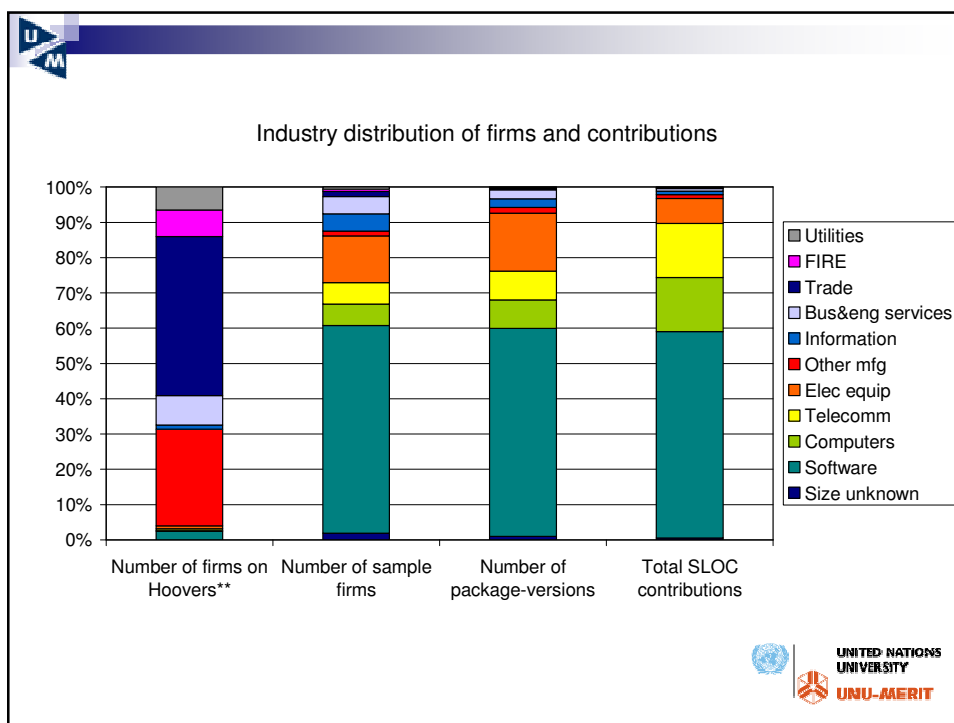
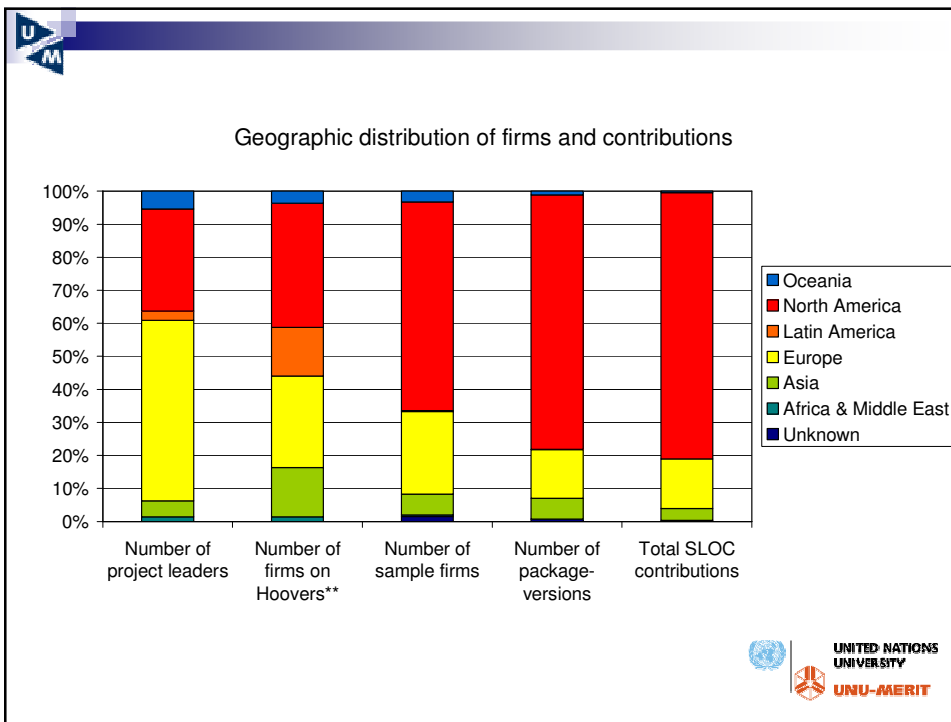
Firm size distribution

Employee size class	Number of sample firms	Share of firm SLOC contribution
unknown*	41	2.3%
Micro: <10	168	9.0%
Small: 10-19	224	6.3%
Small: 20-99	246	11.1%
Medium: 100-499	92	5.9%
Large: >500	211	65.4%
All	982	100.0%



Size distribution of firms and contributions







Probability that a firm contributes to Debian

- Logit regression on grouped data (weighted and unweighted) – 108 observations
- Relative to large North American software firms:
 - Equally likely: European and Oceania firms, electrical and electronics firms, micro-sized firms (perhaps)
 - All others less likely, esp. bus/eng services, FIRE, other manufacturing, non-telecomm utilities, small firms with 10-99 employees
- Explanatory power > 0.9



Size of the average contribution to Debian

- Log (average SLOC) regressed on cell characteristics (weighted and unweighted) – 108 observations
- Relative to large North American software firms:
 - Same size: European firms, computer hardware, telecomm services
 - All others have smaller contributions, with the possible exceptions of Asian firms, electric and electronics firms, and other manufacturing
- Explanatory power approximately 0.5
- Effects are large, typically -100% to -200%





Conditional on contributing

- Relative to large North American software firms:
 - By region: size of the contribution is the same, but Asian, European, and Oceanic firms contribute to fewer packages
 - By sector: most contribute to fewer packages and size of contribution is smaller. Some exceptions:
 - computer hardware firms contribute to more packages and the sizes if their contribution are larger
 - telecomm services contributes about the same as software
 - elec eq contributes to fewer packages but size of contribution the same as software
 - By size: smaller firms make smaller and fewer contributions
- Effects are large, typically -100% to -200%



Summary

- Regressions plus informal evidence show that
 - Largely a US/Canadian/Australian/European activity
 - Concentrated in computer hardware/software
 - Largest code contributors are SW/HW firms like IBM and Sun, also firms exiting the sector (Dec, Netscape, etc.)
 - Many contributors are small hardware firms supplying drivers or other interface software





Future work

- More focus on entry and timing
- Choose a matched sample of non-contributing firms and compare
- Add the patenting activity of these firms – how does OSS interact with IP?



Share of code reuse by type of contributor

