The State of Open Source Server Software

Klaus Ackermann and Shane Greenstein

September, 2018

Comments welcome

Key words: Internet Policy, Internet Services, Open Source, World Wide Web, Global Diffusion

JEL: L860, O330, O340,

Abstract

The study assembles new data to construct a census of worldwide web server use across the globe. We compute market shares across countries and analyze the determinants. We document a large concentration of investment in the United States, and a wide dispersion across scores of countries. We find tens of billions of dollars of unmeasured value in the open source servers. The statistical models show the quality of the country's network and the country's technical sophistication are associated with more web servers, and the innovative environment also plays a role. We find less of a role for economic development, property rights and the rule of law. The findings stress that policies for local supply depend critically on advanced networking and a technically sophisticated populace. While the findings highlight the danger for misattribution in growth accounting, the statistical model points towards the potential to proxy for unmeasured servers with statistical methods.

Department for Econometrics and Business Statistics, Monash Business School, Monash University, and Professor, Harvard Business School, respectively. Corresponding author: sgreenstein@hbs.edu. We thank Susan Athey, Karim Lakhani, Frank Nagle, and Scott Stern for comments. Patrick Clapp provided excellent assistance putting together the exogenous variables. Mercedes Delgado generously gave us assistance with the Global Competitiveness Index. We are responsible for all remaining errors.

I. Introduction

Most research about open source and internet software focuses on production. Much less effort goes to understanding deployment and use. In part this is due to the lower visibility and dispersion of users, and the challenges of collecting systematic information about who or where they are. For such reasons the operations and use of both open source and proprietary web servers remain invisible to all except a small circle of engineers who regularly touch the internet's global operations. This study seeks to address this lack of visibility by assembling the first ever census of global webserver software deployment and by analyzing the determinants of its location across countries.

This study exploits the recent creation of methods for locating longitude and latitudes for devices with internet protocols and uses these methods to count all outward facing web servers in every country on the planet. For reasons explained in the study, we focus on 2012 because it is the earliest year in which it is possible to assemble such data. Due to the data's novelty, our research goals are descriptive and, relatedly, our study seeks to analyze the basic patterns we observe. We analyze the deployment of the three most prominent server software platforms, Apache, IIS, and Ngnix. The largest one, Apache, is believed to be the second most popular open source software in the globe (behind Linux). The second most popular web server is IIS, a proprietary platform from Microsoft. The third server, Ngnix, is the youngest open source web server to achieve large-scale use.

The study initially constructs market shares for each country and describes the patterns. We find that virtually every country in the world contains some use and investment in web servers. Apache, which descended from academic roots, is the largest and most widely used, with more than 15 million copies around the globe. That compares with just under 7 million for IIS and 2 million for Ngnix, which totals more than 24 million servers from the top three platforms. More novel, we find that deployment is quite skewed. The US remains the largest provider of web servers for the global internet. The number of web servers within US borders is 44% of the total and more than six times larger than the levels in the next largest concentration, which is found in China. We also find web servers skew across countries, involve tens of billions of dollars of value, and are not simply proportional to population. We show the skewness correlates with economic development, with the least developed countries displaying greater inequality, as measured by a gini.

The above description of the dispersion of global webservers motivates the three research questions of this study. Why do some countries contain more or fewer servers? Why are there more or fewer per capita servers in some countries? Why does the open source software share vary across countries? Based on a summary of known features of the server market, and following other research, we propose a statistical framework for analyzing the global level and composition of server use across countries. The framework hypothesizes distinct determinants from development, the quality of the networking, the sophistication of human capital, and the institutional environment. Using standard econometric methods for a cross-section of countries, we propose measures of these determinants and assess the relative importance of them.

The findings illuminate the global organization of web servers. While the study finds the primary determinants are the quality of the country's network and the country's technical sophistication, along with the innovative environment. We also find a role for technical sophistication in shaping the share of open source, while we find only a little evidence of influences from other factors, such as the rule of law or property rights. To rephrase, fewer web servers arise when a country contains lower quality networks and less technical sophistication. It is not just income or institutions, per say, that divides provision of the world's internet content.

The analysis has implications for policies encouraging growth of the internet and information infrastructure. We develop some of these, particularly those relevant to local content creation and distribution. Our findings suggest that policies for encouraging local investment in content should build on policies that encourage networks sophistication and technical sophistication in the labor market. Importantly, these are distinct from policies related to software piracy and broad economic development.

I.1. Contribution

The study contributes to the small number of academic studies that expand the data documenting the regional spread and composition of global IT infrastructure (see e.g., Ackermann et al, 2017, Athey and Stern, 2014, OECD, 2013). Like Athey and Stern, the primary effort and novelty involves the assembly of the endogenous variable, in our case, web servers (compared with operating systems in theirs). And like their study, our data are novel in comparison to the best known alternative information about servers, which comes from Netcraft, a private consultancy. Netcraft publishes general results, but it does not make any microdata available for analysis, nor does it provide transparency about its

methods. The absence of public data prevents any analysis of geographic variance. Without such data informed observers cannot discern basic patterns about the global patterns in web server use, such as why some countries have more than others, and why open source is more common in some countries.

We also contribute to the understanding of the deployment of proprietary and open source software. Most analysis focuses on their production and distribution, or analyzes how open and proprietary compete (e.g., Muciano-Goroff, 2018, Wen et al, 2016, Fershtman and Gandal, 2011, Lakhani and von Hippel, 2003, Lerner and Shankerman, 2010, DiBona et al, 1999). When research focuses on the use of open source, research focuses on the productivity of big users and/or contributors (e.g., Nagle, 2017, 2018). To our knowledge, no research focuses on understanding global dispersion and no research performs a global census. In that sense two studies are closest to our study. One is the book by Lerner and Schankerman, 2010, who examine the factors shaping the mix of open source and proprietary code. They analyze a mix of software with surveys from fifteen countries, all of which are developed economies. While our agendas overlap, in comparison we narrow the focus to only one type of software. That comparative narrowness comes with a benefit. It enables us to compare deployment of the same software across the entire globe, which includes many countries from a range of states of economic development. In addition to a more complete census, it enables us to perform basic statistical analysis that had not been impossible.

This study builds on a second paper, Greenstein and Nagle, 2014, which focuses on "digital dark matter." Digital dark matter are inputs into production that otherwise go unmeasured because the activity generates no revenue, and the paper demonstrates a method for estimating value of unmeasured web servers in the United States. Consistent with this prior research, we find that unmeasured webserver software is numerous and valuable. More novel, we show it has spread all over the globe and the unmeasured value reaches tens of billions of dollars. In addition, we also analyze its determinants, which the previous study did not address. Also consistent with prior research, the findings highlight the danger for misattribution in growth accounting. Unlike prior work, here we estimate a statistical model for the presence of web servers, which indicates the potential to proxy for unmeasured servers with statistical methods.

While a large literature on internet infrastructure describes the many symptoms of its growth and spread in the developed and developing world (OECD, 2103), there remain large gaps in understanding the determinants. Relatedly, there are many policies for encouraging local content, which necessarily means encouraging local technical support for it with web servers. Without making

appropriate measurement much of this policy is essentially "flying blind." Reiterating a previous sentence, our study makes a contribution by narrowing the focus in order to enable comparisons between countries across the globe. Due to the novelty of the lens, we begin with basic questions and compare determinants across different types of software. Broadly, this focus builds on prior insight and expands it. Does web server software arise primarily from economic determinants or from institutional ones (Athey and Stern, 2014)? More concretely, does it resemble some IT infrastructure, which largely results from private investment decisions, or does it resemble piracy of operating systems, a product of the institutional environment in a country? While we echo the literature in finding an important role for the network and technical sophistication of the country, our findings about the (small) role of the institutional environment differs from existing research on piracy.

Finally, this paper informs discussions about the spillover from the public subsidies for, and privatization and commercialization of, the internet. As is well known, the internet was invented with subsidies from the Department of Defense, and transferred to the National Science Foundation to support research. It gained wide non-academic acceptance after the NSF backbone shared its infrastructure through privatization. The infrastructure supported deployment of many applications of the World Wide Web, which fueled demand, initially with particular strength in the US (Greenstein, 2015). While the diffusion in the US has been quantified and documented, there remain challenges documenting the diffusion across the globe (OECD, 2013). This study adds an additional dimension for measuring the diffusion of the internet, demonstrating how to measure webservers across the globe. Our findings add yet another piece of evidence that the public costs from creating this technology and fostering its spread were far lower than the gains, and by orders of magnitude.

The paper is organized as follows. Section II explains the basic economics of web servers. Section III explains data and presents descriptive results. Section IV develops hypotheses for the estimation. Section V presents the analysis of econometric determinants. Section VI discusses implications.

II. The Economics of Web Servers

Available descriptive statistics reflect the enormous scale of the internet in the year of this study, 2012. The internet supported 2.4 billion users across the globe and over 634 million web sites.[1] Over time the internet expanded its base of users by increasing extensions to its infrastructure – e.g., more and faster broadband, CDNs, wireless antennae and clients, multiple hardware platforms, and better infrastructure software. In spite of achieving such scale, the basic architecture underlying the network retains much of the same architecture from its initial commercialization. Packets carrying IP addresses travel between switches, servers, and client computers in a format compatible with TCP/IP. Servers and browsers compatible with the World Wide Web feed and interpret the content for users. While elements of "serverless architectures" had begun to emerge, it was a small part of use in 2012.[2]

Web servers are essential for the operations of the World Wide Web. Their primary function is to store, process and deliver web pages to clients. Hence, the content of the internet originates at, or spends time on, web servers. Client and server communicate using HTTP, and frequently exchange files written in HTML, which provides formatting and style sheets for displaying content. Outward facing servers, which this study measures, do not sit behind a security firewall, and do respond to requests for content from any browser. These servers deliver content to any browser requesting it from anywhere on the globe when those requests are made in a compatible format.

That is useful for this study. We count web servers because outward facing web servers must make themselves known to any electronic inquiry coming from a browser. Since three different types dominate our data, we will discuss how they are used, their origins, their similarities and differences, and how these shape our hypotheses and our measurement strategy.

This study focuses on the top three web servers for good reason. The leading private information provider about server use, Netcraft, reports that 86% of all servers in 2012 are either

[1] https://news.netcraft.com/archives/2012/12/04/december-2012-web-server-survey.html, Accessed August 2018

[2] Serverless architectures seek fast delivery and, in some cases, low cost delivery, by avoiding computation and accessing original sources of content. Instead, these architectures leave content on cloud-based storage, where intermediaries provide support. See e.g., https://www.troyhunt.com/serverless-to-the-max-doing-big-things-for-small-dollars-with-cloudflare-workers-and-azure-functions/.

Apache (56%), Microsoft (18%), or Nginx (12%). It reports that 14% come from other sources. It does not break out the market share for any of the other servers, except Google.[3] Google has a 4% market share.[4]

II.1. Origins of Apache[5]

Apache is the most widely used web server. It descended from software invented at the NCSA at the University of Illinois, which also was the home of the creation of the Mosaic browser, the first browser to gain mass market acceptance. Apache's ancestor was called the NCSA HTTPd server. This was the most widely used HTTP (Hypertext Transfer Protocol) server software in the research-oriented non-commercial internet. The server was a collection of technologies that supported browsing and use of Web technologies.

The HTTPd server software followed an unexpected path into widespread use. The server software first became available for use as shareware in the early 1990s, with the underlying code available to anyone, without restriction. Many Webmasters took advantage of the shareware by adding improvements as needed or by communicating with the lead programmer, Robert McCool. McCool, however, left the University (along with others) to found Netscape in the spring of 1994, and thereafter the University was slow to replace the coordinator. By early 1995 there were eight distinct versions of the server in widespread use, each with some improvements that the others did not include. These eight teams sought to coordinate further improvements. They combined their efforts, making it easier to share resources and improvements, allowing them to build further improvements on top of the (unified) software. The combination of eight versions became known as Apache (ostensibly because it was "a patchy web server"), and, informally at first and more formally over time, the group adopted the practices of open source. Apache had become so widely used – by the time the University appointed a new coordinator – the university saw no point in further supporting HTTPd server, and abandoned its role as coordinator, passing it to the Apache organization.

[3] https://news.netcraft.com/archives/2018/05/29/may-2018-web-server-survey.html. It does not report its methodology for how it came up with this estimate. Accessed June, 2018.
[4] Many of these are forks of open source projects, where the owner has adapted and optimized the server software to specific uses. The most well-known of these is operated by Google in its data centers, and there are many others, some for sale, some not. In many cases, the servers do not make it easy to learn much about them.
[5] Much of this section draws on Nagle and Greenstein, 2014.

The transfer never involved any licensing or monetary transactions, and has never had a price affiliated with either its inputs or outputs. Its growth and deployment has largely taken place without standard economic measurement.[6]

The lack of monetary transactions became embedded in many aspects of Apache. Its sponsoring organization relied upon donations and a community of technically skilled users who provided new features for free, motivated both by the intrinsic and extrinsic rewards. As with other open source software (Lakhani and von Hippel, 2003, Lerner and Shankerman, 2010), Apache eschews standard marketing/sales activities, instead relying on word-of-mouth and other non-priced communication online. Apache also does not develop large support and maintenance arms for their software, although users do offer free assistance to each other via mailing lists and discussion boards. At most, Apache is affiliated with revenue-generating activity that are complementary to its use, such as a large labor market for Apache programmers, administrators, and third party consultants.

Apache grew in popularity as the commercial internet grew, becoming widely used in the customer facing and procurement activities of many firms. It is regarded as the second most popular open source project used by businesses, after Linux.

II.2. Origins of IIS

Microsoft offers a web server called IIS (i.e., internet Information Services). Its origins date back to Microsoft's entry into providing software to support the World Wide Web as part of Microsoft's efforts to provide server software in Windows NT. As part of the Microsoft family of products, it plays a role in a larger suite of revenue-generating activities. Considerable debate among web masters surrounded the merits of the earliest versions – whether they contained artificial limits on supporting multiple web pages, and whether contracting disadvantaged competitive servers. All analysts agree that it became widely used as it developed new functionality, and, just as with Apache, IIS developed a large labor market for programmers, administrators, and third party consultants.

IIS is the only prominent proprietary web server software for sale to others. Over time it has come in a variety of formats, editions, and prices. At the time we observe it, Microsoft puts its sale and

[6] This path notably differed from the browser, which diffused into general use through two channels. A team of programmers from the University founded a firm, Netscape, which pioneered the commercial browser market. The University of Illinois also licensed the Mosaic browser for millions of dollars. Through such a licensing deal, the browser reached Microsoft, who used it as the starting point for designing Internet explorer.

support behind the product, which fostered its use across many industries and geographies. IIS's users say it possesses appealing features, including its compatibility with other Microsoft products, as well as its certification, documentation, and ease-of-use in enterprises with routine requirements.

There was considerable rivalry between Apache and IIS, which engendered debate among web masters about which situations best suit which option among the web servers. Microsoft benefited from suspicion among some large organizations about using open source code they had not vetted. It also was hurt by the strong desire among many web masters to retain autonomy to modify software, as well as general enmity from some technical communities towards the firm.

Here is where we leave it. Sometimes IIS made more sense for a user than Apache, and vice versa. We do not need to resolve this debate to analyze the determination of use of servers.

II.3. Origins of Nginx

Nginx is a late entrant in comparison to Apache and IIS. It has a unique set of origins. Programmer Igor Sysoev started the initial work in 2002 when he sought to scale the server for a large online media company, optimizing it to handle at least 10,000 of concurrent connections. In 2004, on the 47th anniversary of Sputnik[7], Sysoev opened Nginx to the public as an open source project, using a BSD license for open source. Steady improvement from many contributors turned the software into a viable web server around 2007.[8]

Nginx's origins play a role in its functional appeal. It performs well on benchmarks that stress large volumes of traffic, and that performance gave it a foothold in media and entertainment enterprises with high peak loads. That achievement came at the cost of sacrificing some of the adaptability found in Apache, which – oversimplifying a long technical explanation for the sake of brevity – had come closer to adopting a "one-size-fits-all" approach to its design. It also gave Nginx a different appeal than the ease-of-management of IIS, which came with considerable support. This gave nginx a foothold from which to grow, and over time the software community around Nginx added extensions and modifications in an attempt to grow out of this niche.

[7] https://www.nginx.com/blog/nginx-vs-apache-our-view/, accessed June, 2018.
[8] This is Sysoev's own estimation, see http://freesoftwaremagazine.com/articles/interview_igor_sysoev_author_apaches_competitor_nginx/, accessed June, 2018.

In July, 2011, a company was founded by Sysoev and others, and it is also called Nginx (see nginx.org or nginx.com). It is based in San Francisco, and Sysoev serves as CTO. At the time of our data, this company supports “Ngnix plus”, which includes enterprise level services. It offers a set of paid extensions on top of the open source Nginx, which also continues to improve. These commercial extensions target long time users who desire commercial-grade features that are not normally available in any existing open source product. It also targets enterprises which require both technical support and license payments. These help its use as an "edge web server" for the cloud, hosting and CDN service providers.

Due to the timing of the founding of the for-profit organization, we expect the vast majority of the measured Nginx in 2012 did not generate revenue. We will treat it as open source.

II.4. Implications

In most settings a web master sets up a web server for continuous operation, and servers are treated like any investment with a onetime setup cost and a regular maintenance cost. While it is difficult to make general statements about these fixed and variable costs, it is useful to identify a few common patterns of investment and use. That will inform the framework for analyzing webserver use.

Web servers are comprised of both hardware and software, as well as complementary investments. The range of complements is extensive: electricity supply and support, registration with the key governance of internet addresses, physical connections to equipment to perform data traffic functions, and a range of tools to manage traffic loads. In short, web servers require technical skill to install and maintain, and the labor costs of a large installation also can become substantial.

The setup costs can be comparatively high at small scale, and beyond that there is considerable debate. To an expert the incremental costs of maintaining the next server declines with increasing scale, especially if the servers are part of a rack of similar servers. There is considerable debate among web masters around the operational trade-offs between using fewer or more servers for a variety of content, or to manage a large load of content with different peaks in demand. While per unit costs of setup and operations are thought to decline with routinization, which increases with scale, performance varies a lot between different types of content and applications – e.g., static web pages, time-intensive news, and streaming video. These complexities prevent simple characterization of the trade-offs, and, relatedly, prevent simple characterization of the fixed and variable costs of supply.

In developed countries all the inputs are widely available, and at competitive prices in any major urban center with electricity and internet connections, a thick supply of technical labor, and providers of services. All prominent web servers have large online and offline communities behind them, with Nginx among the newest. There are also large markets for the appropriate equipment to support, maintain, and upgrade them, as well as skilled labor to manage them. In developing economies the situation can be more challenging for users. Complementary inputs, such as internet connectivity or reliable continuous supply of electricity, may not be available at the desired quality in the desired location. Other inputs and skilled labor markets may also be lacking. The institutional environment may also discourage long term investments required of a network with such distributed investments and responsibilities.

The above discussion guides statistical analysis. The costs of operating Web servers to deliver content does not fully determine the use of it, and does not entirely determine which attributes deserve the most attention. Those costs will vary across the globe, and will be higher in some countries than others. Those costs will shape the supply of content by encouraging or discouraging investment in web servers. We should expect the level of investment in servers and the levels per capita to vary across countries. To the extent such costs differ between open source and proprietary software, we also expect the market share among proprietary and open source shares to vary.

II.5. A Framework

Two benchmarks help motivate different frameworks for the endogenous variables. Define one benchmark as "content travels to any user on the globe." In an internet with worldwide distribution and supply of content, the location of a web servers in a country provides information about the origins of the supply of content for the global internet. In this approach, a country's share of global server market provides information about that country's supply of content for the global internet.

Another benchmark motivates a different endogenous variable. Define it as "content travels to only users within the nation." In that situation, one country's servers serve only a local population and no other. The location of the web servers in a country provides information about the supply of content for a country's population. In this approach, the level of per-capita servers provides information about the country's supply of content for its own population.

The first benchmark focuses on market share *between countries* in the global internet, i.e., variance in the relative levels of investment compared across countries. The latter focuses on

differences *between countries in their* servers per capita, i.e., variance in the relative levels of per capita investment. In practice, of course, the outcome lies between these benchmarks, so both deserve attention.[9]

That discussion leads to the third endogenous variable. In either framework, each country's webservers face local costs, and installers have options among at least three prominent web servers, where two are open source and one is proprietary. If the differences in costs and benefits between proprietary and open source affect user decisions – say, because one is more costly than the other – then the market share for proprietary and open source software will be sensitive to features of a country that nurture its use. That focuses our attention on the market share *within a country* of proprietary servers (Microsoft) instead of open source (Apache and Ngnix). In other words, do differences in shares of open source software follow predictable patterns?

The above discussion suggests four key factors determine the level, per-capita level, and open source share of webservers:

- **Economic development.** Web servers follow internet activity, and supports operations and communication. Growth in internet activity follows growth in general economic activity. Servers may be sensitive to the level of economic development of a country.
- **Quality of the network.** Web servers must work with complementary functions in order to deliver services. Placement of servers may be sensitive to the extent and quality of investment in other networking in the local area and within the country.
- **Technical sophistication.** Web servers require technical skill to set up and operate. Servers may be sensitive to the technical sophistication of the local labor market in a country.
- **Institutional environment.** Many firms treat servers like many other long term investment. Servers may be sensitive to the same institutional and environmental factors that shape other technical and inventive investments in a country.

We will divide the analysis of the causes of variance into four categories. Which factors determine the regional composition of web servers? This is the open question we address next.

[9] Existing evidence suggests there are large world wide data flows, but much less about the extent of "trade" in digital goods across borders. See, e.g., Blum and Goldfarb, 2006, for the first attempt at this question with early data from the young commercial internet, and Aguiar and Waldfogel, 2014, for a more recent examination of trade in digital music, or Gomez-Herrera and Martins, 2014. All support the view that content supports a mix of local and cross-border demand, and home bias persists, suggesting that actual practices mix the two benchmarks.

III. **Data Description**

The data for this study comes from a computer virus called "Carna Botnet", which infected 420 thousand computers worldwide to completely port scan the whole internet in 2012 (Carna 2013). The author remained anonymous, but the data and information collected became public domain for others to analyze. We follow the method outlined in Ackermann and Angus (2014) for extracting the service probe information from this data source. It is a common norm of the internet for webservers to use port 80. The computer virus sent a TCP GetRequest to port 80 and recorded the response. An example response would be a string like:

HTTP/1.1=20200=20OK=0D=0AContent-Type:=20text/html=0D=0ALast-Modified:=20Thu,=2020=20Oct=202011=2014:59:18=20GMT=0D=0AAccept-Ranges:=20bytes=0D=0AETag:=20"3fcdcdd7388fcc1:0"=0D=0AServer:=20Microsoft-IIS/7.5=0D=0AX-Powered-By:=20ASP.NET=0D=0ADate:=20Fri,=2014=20Dec=202012=2021:44:38=20GMT=0D=0AConnection:=20close=0D=0AContent-Length:=20689=0D=0A=0D=0A<!DOCTYPE=20html=20PUBLIC=20"-//W3C//DTD=20XHTML=201.0=20Strict//EN"=20"http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">=0D=0A<html=20xmlns=3D"http://www.w3.org/1999/xhtml">=0D=0A<head>=0D=0A<meta=20http-equiv=3D"Content-Type"=20content=3D"text/html;=20charset=3Diso-8859-1"=20/>=0D=0A<title>IIS7</title>=0D=0A<style=20type=3D"text/css">=0D=0A<!--=0D=0Abody=20{=0D=0A=09color:#000000;=0D=0A=09background-color:#B3B3B3;=0D=0A=09margin:0;=0D=0A}=0D=0A=0D=0A#container=20{=0D=0A=09margin-left:auto;=0D=0A=09margin-right:auto;=0D=0A=09text-align:center;=0D=0A=09}=0D=0A=0D=0Aa=20img=20{=0D=0A=09border:none;=0D=0A}=0D=0A=0D=0A-->=0D=0A</style>=0D=0A</head>=0D=0A<body>=0D=0A<div=20id=3D"container">=0D=0A<a=20href=3D"http://go.microsoft.com/fwlink/?linkid=3D66138&clcid=3D0x409"><img=20src=3D"welcome.png"=20alt=3D"IIS7"=20width=3D"571"=20height=3D"411"=20/>=0D=0A =0D=0A</body>=0D=0A</html>

This is similar to the first response a web browser would receive when accessing the server. The string can then be passed to the scripting engine of nmap to convert this information into a tuple of (http,Microsoft IIS,httpd,7.5, Windows). We only retain the fact that the IP address the probe was sent to is a Microsoft IIS server and ignore the different version numbers. As the port scan was run more than once, we aggregate the information to unique IP address level, avoiding the issue of double counting of the web servers. Each individual IP address is then joined based on the timestamp to the closest revision of an IP geo-location database (Ackermann et al, 2017). The geo location database is from Digital Element, a commercial company with high accuracy (see e.g., Gharaibeh et al, 2017). The individual probes are then aggregated to a country.

Our definition of country follows international norms. We use the ITU definition of a country as default, which corresponds to the ISO 3166 standard.[10] We exclude disputed territories (e.g., Kosovo), where it is not possible to obtain reliable data. This exclusion affects a tiny number of servers.

III.1. Levels and Location

[10] See, e.g., http://www.statoids.com/wab.html. Accessed June, 2018.

We count the number of servers in each country. This information covers the location of over 26 million web servers in 253 countries. We focus on the three prominent server software platforms and exclude servers that have been modified for custom purposes, such as Google's or Facebook's servers. Each country's count is divided into three categories for Apache, Nginx, and IIS. We do not count any server if we cannot associate it with one of these three.

Aggregating a count to the level of a country has many advantages. It provides a natural definition of open source and proprietary at the country level. In the analysis below, Microsoft's share of servers represent the share of proprietary software. In addition, it is robust to measurement error. Though we do not expect many errors in location, the use of broad country borders reduces the measurement errors. If error is randomly distributed across geography, there is no reason to expect it to bias inferences based on cross-country comparison.

These counts also have several drawbacks. They are not weighted by size or vintage. That should not matter if size and vintage are randomly distributed. This seems plausible when comparing Apache and IIS. However, this seems doubtful in the case of Nginx, which has a built-in bias towards a younger vintage of installations and high volume of transactions. While one typical Nginx server probably supports more activity than one average Apache or IIS server in the sample, there is not a natural weight for them. We do not have an approach to dealing with this, so we leave this topic for future research.

Figure 1 provides a visual representation, i.e., a "map" of the distribution of global webservers. Servers are generally found in population centers in the US. It generally locates in similar places across globe, except in Africa. The US and Europe have "densest" provision. Cities in Asia – Japan, China, India – also visually standout. This visual representation does not help understand differences in per capita or market shares, so we next turn to statistical descriptions of these patterns.

Table 1 provides a description of the location for the top thirty countries with the largest concentration of 24,282,089 million servers worldwide.[11] The US is by far the largest provider of web servers, providing 44% of Apache, 46% of IIS, and 27.6% of Nginx. It is just under half of the world's supply and more than six times larger than the next largest supplier, which is China. The next biggest suppliers are largely unsurprising – China, Germany, Japan, UK, Netherlands, Canada, France, Russia, and

[11] This table excludes Romania, which contains an inexplicable and implausible miscount of Nginx servers. This is the only medium size country excluded from the data. It reported 305k Apache servers, 20k IIS servers, and 2.1m IIS servers. The first two are small percentages of worldwide use, while the latter would be more than half of reported Nginx use, which is implausible. This is such an outlier that we must exclude it from statistical analysis.

Brazil. It takes the sum of the servers in the next twenty countries to get to the same amount as found in the US. In short, the global distribution is quite skewed.

China has an overrepresentation of Microsoft server, which might be a consequence of the regulatory framework in China. Companies are required by law to abide the censorship rules, demanding that any website that allows discussion needs to filter it through the automated censorship software (King et all, 2014).

The share of servers otherwise has unsurprising features. Apache has highest overall market share, Microsoft is second, and Nginx is a respectable third (and, by public reports, growing faster than the other two at the time of this sample[12]). Together the open source share is 71.5% of total. We note an interesting difference with other public sources. Open source is 2.5 times the size of proprietary software. This is lower than what is reported (79%) by Netcraft, whose numbers need adjustment to make an apples-to-apples comparison about Apache, IIS, and Nginx.[13] Netcraft reports that open source is 3.8 times the size of proprietary software. In other words, this dataset finds comparatively less open source software than the only other publically known source of information.

For assessing the unequal distribution of the webserver worldwide, we calculate gini coefficient on per country and per capita bases. The world gini coefficient is 0.78, where a value closer to 1 means more inequality across countries. Separating the countries into the corresponding world development income brackets, a more nuanced picture emerges. The values range from a gini of 0.92 for lower middle-income countries, a value of 0.68 for upper middle-income countries to 0.48 for high-income countries. In short, the higher the income group the more equal the servers are distributed among them. Low-income countries were excluded as they only account for 0.1% of the total webserver share. Figure 2 visualizes the distribution. These findings provide part of the motivation for latter analysis: Why do some countries have more servers than others? Is this merely a function of economic development?

We also compared gini coefficient within countries, using the latitude and longitude locations as unit of observations and the number of servers at a location as measure of interest. Overall, if we filter by at least 50 locations by country, gini coefficients range from 0.83 (Denmark) to 0.97 (Ireland), hinting at a high concentration to a few locations within the countries, probably in major cities. These findings

[12] According to Netcraft's survey of web services. See https://news.netcraft.com/archives/2018/05/29/may-2018-web-server-survey.html. Accessed June, 2018.

[13] We adjusted Netcraft's reported estimates to account for only the top three platforms. The comparable number for open source servers is (56+12)/86 = 79.0%. Netcraft does not report its methods, so we do not know how to reconcile the difference between their estimate and ours.

suggests the additional possibilities for measuring the spread of webservers. Pursuing this topic would take us astray from this study's research goals, and we view finer geographic measures as an avenue for new research.

III.2. Value of Servers

What is the value of the web servers on the global internet? The answer is unknown because so much of it is digital dark matter. We investigate this question to further motivate the study. Is this component of the internet large enough to be worthy of interest? We will answer yes.

To make an estimate, we use the same method as in Greenstein and Nagle (2014), which puts a monetary value on the Apache HTTP Servers and Nginx HTTP servers by comparing it with the most widely used proprietary and pecuniary choice. This approach follows Nordhaus (2006), who states that (p. 146) "the price of market and nonmarket goods and services should be imputed on the basis of the comparable market goods and services," and (p. 151) valuation "should rely on available market and behavioral data wherever and whenever possible."

In 2012 the most prevalent web server is Microsoft's IIS. IIS is shipped for free with Microsoft's Windows Server 2008 operating system, the price of which varies greatly. In 2012, the price for Windows Server 2008R2 Datacenter Edition was $2999 for one license. The most bare-bones version of Windows Server 2008, called the Windows Web Server2008, is priced at $469.[14] This version of Server 2008 is intended purely for "the development and deployment of internet-facing Web sites and services."[15] What is a representative price for IIS? We expect that neither seems particularly informative of the true price users tend to pay. We utilize these two price points to understand the range of possible prices and take its average ($1734) for some ballpark estimates.

A simple calculation suggests the mismeasurement of open source server software is large enough to be economically important. The value of the stock of all servers lies between $11.2B and $72.8B, averaging $42.1B. That puts the value of Nginx at between $0.9B and $5.8B, averaging $3.39B. The value of unmeasured open source falls between $8.1B and $52.1B, averaging $30.1B. Only the IIS numbers makes a contribution to global asset tables and GDP calculations (but only when a transaction

[14] This falls somewhere between the prices for Windows Server 2008 R2Standard, which is $1029 for five licenses, and Windows Server 2008 R2 Enterprise is $3999 for twenty-five licenses.

[15] Finally, IIS also comes installed with Windows 7, which can be purchased for as low as $119.99. However, Windows 7 is not designed to be used as a production scale web server and it is unlikely that any company hosting a public website would use this version of Windows.

occurs), while the open source value is digital dark matter. Summarized at face value, the percentage of open source servers (71.5%) and the total value of servers ($42.1) supports the conclusion that a large fraction of an important asset for the global internet goes largely unmeasured.

The last two columns of Table 1 displays the proprietary share and value of web servers for the top thirty countries. While the US has the largest level of unmeasured assets, the table shows that the number could be large in many countries. The dispersion of investment in servers in Table 1 reinforces the research questions. Why do some countries have more open source software?

III.3. Statistical sample

For econometric analysis we will examine 213 countries. These are the 213 largest countries in the sample, with a few exceptions (such as Liechtenstein). We got from 253 to 213 by dropping countries for whom there is scarce information about economic activity and the institutional environment, and who play tiny roles in the global internet. Most dropped countries are islands and small territories.[16]

Table 2a presents the list of the 213 included countries. The modal country in this dataset is a middle to low income country. Table 2b presents the general statistics for the 213 countries. One notable feature is the absence of zero investment. Every country contains some investment in web servers and virtually every country contains investment in one of the three platforms. If a country lacks investment in a server from one of the three, it lacks it in the newest among them, Nginx.

In this sample, the levels of Microsoft and Apache are highly correlated (0.96), and so are their logged levels (0.95). The level of Nginx is weakly positively correlated with Microsoft (0.29) and Apache (0.29), and the logs more so (respectively, 0.85 and 0.89). In other words, the level of Apache is easily predicted by Microsoft and vice versa, and the logged level of Nginx is too. This simplifies our analysis at the outset. Instead of analyzing the levels of each platform, we will examine aggregate levels, i.e., total number of servers. Then we will examine per capita, and market shares of proprietary and open source.

[16] The dropped countries include Aland Island, Anguilla, Antarctica, Antilles, Bonaire, Sint Eustatius, and Saba, Bouvet Island, British Indian Ocean Territory, Christmas Island, Cocos (Keeling) Islands, Cook Islands, Falkland Islands, French Guiana, French Southern Territories, Guadeloupe, Guernsey, Heard and Mcdonald Islands, Jersey, Martinique, Mayotte, Montserrat, Niue, Norfolk Island, Pitcairn, RAF Ascension Island, Reunion, Saint Barthelemy, Saint Helena, Saint Pierre and Miquelon, South Georgia and South Sandwich Islands, Svalbard and Jan Mayen Islands, Tokelau, US Minor Outlying Islands, Vatican, Wallis and Futuna Islands, and Western Sahara. In addition to accounting for a tiny number of servers, these countries account for a tiny fraction of world GDP.

The first three rows of Table 2b show a highly skewed variable, which is always positive, except in a few instances of Nginx, which are zero for countably small number of countries. We will need to account for this distributional feature of servers. That will necessitate using log transformations.

IV. **Hypotheses and Measurement**

What factors shape adoption of server software across countries? Building directly off the foregoing discussion, we describe the list of variables and predictions. Table 3 provides a summary.

Economics development: A country's income could play a role in server investment. It should operate through both demand and supply. High income creates demand for internet services and low income reduces it. Economic development also correlates with budgets for private investment and, relatedly, for budgets to operate many servers on a regular basis. We expect income to shape the investment behavior of firms, with higher income countries containing more capital-augmented IT than lower income countries.

We expect income to have a different effect on the share of open source. If income is lower, we forecast that web masters will face more budget constraints. They may try to "save money" or "substitute webmaster time for money" by using open source software. Therefore, we predict, ceteris peribas, lower income settings will be more likely to have a high fraction of open source software.

We proxy economic development with two variables commonly used in cross-country comparisons, the log of the country's per-capita income in 2012,[17] and the percentage of the population with electricity in urban areas in 2012 (or 2011 when the former is not available).[18] Per-capita income is skewed, so we log it. Both of these vary with economic development and capture something distinct about the level of development.[19] The latter captures differences in quality of public infrastructure.[20]

[17] According to the Worldbank Development Indicators, accessed April, 2017.
[18] We take percent for urban areas instead of for the entire country, because the former is more widely available. See Worldbank Development Indicators, accessed April, 2017.
[19] As with any cross-country regression we cannot merely proxy income by GDP levels, which is highly endogenous with the level of country-wide investment and, in this case, networking. We follow the literature, and find a scale-free measure, such as gdp per capita.
[20] Following Athey and Stern (2014), we also experimented with other factors predict capital investment, such as the inflation rate for a country. We never found evidence that inflation rates shaped any decision in 2012. With a premium on a small number of explanatory variables, we dropped the variable.

Network Quality: Servers are complementary with many other inputs that together comprise the supply chain for Internet services and activities. Better complementary networks inputs lead to better server performance, so we expect better networks to support a higher level server software investment, ceteris paribas. It is less clear how higher quality networks shape the share of open source or proprietary software. Both types of software ought to be sensitive to network quality, so we have no prediction about the relationship between network quality and open source software share.

We proxy for network quality with two measures, the log of the average broadband speed and the log of broadband price in the country.[21] Both vary with level of development. Nonetheless, there is considerable variation in each, related to a range of government programs and local market factors.

Technical sophistication: A minimal level of technical sophistication is required for use of any software. So all web server software should increase with a more sophisticated labor market. We expect this to matter most for open source software, which tends to lack the corporate support that eases use. The level and share of open source software may increase with technical sophistication.

We seek a proxy for technical sophistication of the country that is not correlated with other measures of network quality. One such measure is patents per capita.[22] There is discontinuity in this measure at zero, because many developing countries have no patents.[23] We include this dummy for these countries with low technical sophistication. We log non-zero patents and make it zero otherwise.

Institutional environment: The institutional environment should shape investment. The property rights theory of investment predicts that better property rights protects private incentivizes for more investment. If there is pervasive large scale evasion of property rights and lack of enforcement of property rights, then it may discourage any investment in server software. Thus, we expect the environment to predict the level of investment in server software. Better enforced property rights also should contribute to less piracy, which should contribute to more use of proprietary software. Hence, ceteris paribas, we expect better property rights to decrease the market share for open source software.

There are a variety of ways to measure variance in the environment between countries. To facilitate comparison with Athey and Stern, 2014, we employ measures similar to theirs, which they based on the literature on cross-country differences in the legal environment. The first measure comes

[21] According to the ITU database, accessed April, 2017.
[22] This is a count of US patents registered to assignees in that country.
[23] We observe a maximum of 214 countries, and 68 countries have no US patents.

from the Heritage Foundation's index of property rights.[24] Second, we use index measuring the innovative environment from the Global Competitiveness Institute (Delgado et al, 2012).[25] The third proxy is an index of the Rule of Law, which comes from the World Bank.[26]

Table 4 includes a statistical summary of the exogenous variables. The first column of numbers illustrates the challenge for statistical analysis. Few of the variables are available for all countries. This lack of available data frames a tradeoff in constructing the regressions. We cannot have both a large sample and more variables to describe a country's features.[27] Because observations come at a premium, we will favor more observations and economize on the numbers of exogenous variables.

We face a challenge using the three proxies for the institutional environment, which are unavailable in many countries. Using all three reduces the sample size and renders estimation virtually infeasible. We adopt a strategy of estimating equations first with economic development, network quality, and technical sophistication. Then we added each of the environment variables, estimating each separately.

V. **Results**

V.1. Level of Servers

[24] The Heritage Foundation maintains a variety of measures of property rights, but these tends to be highly correlated in this dataset. Hence, we use the Heritage Foundation Economic Freedom Indicators Index for 2012.
[25] Athey and Stern use the general competitiveness index, while we opt to use innovative index. While these two are highly correlated, the innovation index is better suited to our setting, just as the competitiveness index was to theirs. We thank Mercedes Delgado for giving access to this data. See Delgado et al (2012) for longer explanation.
[26] As with the others, this comes from the World Bank Development Indicators.
[27] In practice availability placed a large constraint on measurement. Other variables for measuring similar concepts were usually rejected either because they were not available, or, within the range they were available, they correlated with the variables presented in Table 4.

What determines the level of investment in servers in some countries and not others? Table 5 displays the results. This shows the result for the total number of servers.[28] The regression is also equivalent to answering: What determines the share of global servers in a country?[29]

Table 5 tells a straightforward story. Income matters when free standing, but loses significance in the presence of measures of the network and technical sophistication. Income determines the quantity of servers, only when the measure of network quality is absent. In the presence of robust measures of the network quality and technical ability of a country, income has no estimated effect. This suggests that income determines a more technically accomplished network, which, in turn, actually determines the size of the servers found in a country.

All measures of network sophistication matter. The elasticities for broadband speed and broadband price are high – i.e., 0.41 and -0.68. These are economically meaningful. A one standard deviation change in speed yields 0.41*1.38 = 0.566 of change in endogenous variable and one standard deviation of change in price yields -0.68*0.87 = -0.592 of change in endogenous variable. Compared with one standard deviation of endogenous variable of 2.9, these together explain a significant fraction of the observed variance in outcomes.

Technical sophistication also matters. Thirty percent of the countries have no patents, which accounts for a large fraction of the difference between extremes, which is more than 14 (min = 1.70 and max = 16.18). In other words, most of the countries with a countably small number of servers are so technical unsophisticated that they do not have even one patent. Once that is accounted for, which puts the data in middle and high income countries, then the elasticity is moderately high. A one standard deviation in patents 0.5*6 = 0.3 accounts for a moderate growth in servers.

A comparatively small number of determinants predicts web server software. Column 2 contains six variables, and this explains 64 percent of the variance across countries. Despite the relatively lack of visibility, this suggests the potential for using statistical methods to forecast levels of servers.

[28] We do not show results for each type of server for the sake of brevity. Preliminary analysis showed nearly identical estimates when the endogenous variables were the logged level of Apache and IIS servers. This is not surprising because the two variables are so highly correlated. The qualitative results for the logged level of Nginx servers was qualitatively interesting, and will receive attention below in the discussion about proprietary share and open source share.

[29] This is because all variables could be rescaled by dividing by total number of servers globally, which would have no effect on any estimate except the constant.

The environmental variables shrink the size of the sample and, as expected, make estimation challenging. The innovation index and the Rule of Law index matter in this specification, with moderate elasticities. They seem to measure distinct effects and one of them predicts in the wrong direction. A one standard deviation in GCI of innovation yields 2.26*0.29 = 0.654, and a one standard deviation increase in the index for the Rule of Law reduces servers -5.56*0.14 = - 0.778. Both are of moderate importance. The latter has no obvious interpretation. While it could be interpreted as consistent with discouraging open source investments, as we show below, later results will reject that interpretation.

At a minimum we conclude that that innovative environment shapes investment in servers by shaping innovation, innovation operates through mechanisms other than property rights or rule of law. Consistent with the findings on technical sophistication, that points towards the role of human capital, such as training, education, and factors improving skilled labor markets.

V.2. Per Capita Servers

Which countries supply more server software relative to the size of their population? If most server content goes to local users, then this measures the strength of local supply of internet capacity. It also could indicate export/import of internet content by those with strong/weak internet infrastructure relative to population levels.

Table 6 shows the results. In this specification the income variables predict investment. GDP per capita is positive and significant in all specifications. It has a moderate effect. The estimate declines once other factors are included, varying between 0.55 and 0.84. A one standard deviation increase in log of GDP per capita would result in 1.5*0.63 = 0.95. That is more than a third of one standard deviation in the endogenous variable, which is 2.58. The coefficient for electricity is also positive and significant in most specifications. However, it has a small economic effect. Even a large coefficient estimate yields small changes from a one standard deviation improvement in electricity (0.025*20=0.5). This factor matters in extreme situations, when a lack of electricity predicts an absence of servers. For example, if small electricity serves only a small fraction of the urban population – such as two standard deviations below the mean, a difference of 41—then the servers per capita will be 0.025*41 = 1.0, which is a large effect.

The measures of network sophistication tell an ambiguous story. The coefficient on broadband speeds is positive and significant in two of four specifications. The coefficient on price is negative and significant in only one specification.

Technical sophistication does matter. Patents per capita has a positive effect, albeit a moderate effect. One standard deviation increase yields small effect 2.73*0.3 = 0.82, which a third of one standard deviation of endogenous variable. No patents has anticipated negative effect and it is large. Consistent with the estimates in the previous table, lack of patenting indicates a difficult situation and poor supply.

Similar to the prior estimates for levels of servers, a comparatively small number of determinants predicts per capita web server software. Column 2 contains six variables, and this explains 80 percent of the variance across countries. Again, this reinforces the conclusion that statistical methods could forecast total per capita server use.

Only the Heritage index of property rights has a statistically significant effect, but it is small, at 0.02. A one standard deviation in the index results in a small change in the endogenous variable, 11.5*0.02 = 0.23. None of the other indices about the environment have an effect. We conclude that the environment largely does not shape the per capita supply of servers. Per-capita supply is much more sensitive to economic and technical forces.

Overall, the two sets of findings together suggest countries with higher incomes do not necessarily support larger supply of servers. It is important to benchmark against population levels. High income does support larger supply of servers in comparison to the size of the local population. That continues to hold even in countries with higher quality network – especially in terms of speed and sometimes price, which support more servers in those countries. However, it has little effect on per capita investment. Finally, in either case, the technical sophistication of a country – as measured by the propensity to invent – supports more server software within their borders and more per capita server investment.

More surprising, little evidence suggests that property rights or enforcement of the law has any effect on the level of servers within a country. The only relevant factor are those that shape the environment for all innovation.

V.3. Open Source

We let the endogenous variable be the proprietary share of software. This is represented by Microsoft share of servers in a country. Open source is, correspondingly, represented by the sum or Apache and IIS servers. Though amounts of proprietary and non-proprietary software are highly correlated, the differential effects of determinants on types of open source and proprietary web servers could lead to different shares of proprietary software in different countries.

The distribution of market share has characteristics that require attention in the statistical analysis. Every country has at least some proprietary and some open source software, so the shares vary between [0,1]. Though the median and mean are not statistically different from one another, there are a few more observations at the extremes than would forecast from a normal distribution. The variable is "centrally distributed with fat finite tails," which violates standards OLS assumptions.

We follow Papke and Woolridge (1996) and take a parametric approach to transforming a variable which is a non-zero proportion and "fat tails." We let Y represent the endogenous variable. Recognize that the logistic function is one functional form with "fat tails." If we begin with the assumption of a logit function, then $\ln[Y/(1-Y)] = \ln(Y) - \ln(1 - Y) = XB$. This transformation has an intuitive appeal since it uses the log of the ratio of proprietary to non-proprietary software.

Table 7 shows the distribution of the endogenous variable and transformed endogenous variable. The transformed endogenous variable is zero when a country has equal shares of proprietary and non-proprietary software. The average country tends to have more open sources software, as shown: the mean level is negative.

Table 8 shows the results of a regression analysis. What increases proprietary share? In most specifications GDP per capita increases the proprietary share of software, particularly in specifications with larger samples of countries. Electricity supply is not a robust effect. This partly follows expectations. The coefficient on GDP per capita is estimated to be a moderate effect, ranging from 0.1 to 0.35. We illustrate on a middle estimate. A one standard deviation change yields 0.26*1.5 = 0.39 change, which is over one third of one standard deviation (0.94) of the transformed endogenous variable. In short, this matters most in comparisons of the poorest and richest countries.

Network quality does not matter. The results for technical sophistication are somewhat mixed. Patents per capita matters in most specifications, ranging from -0.09 to -0.14. As expected, more technical sophistication reduces use of proprietary software. One standard deviation in patents per capita equal to 2.73, so even at highest estimate the effect of one standard deviation increase is 2.73*-0.14 = 0.38. This is a small effect. Meanwhile, in samples with large numbers of countries, variable for no patents has large economic effect. Minimum effect of 1.83 and largest is 2.24, which exceeds twice the standard deviation of the endogenous variable. This result suggests that the least technically capable countries use much less open source, ceteris peribas.

Among the measures of property rights, none of the indices has any predictive power. In other words, open source share is not sensitive to property rights or the rule of law. This is a surprise, and contrasts with piracy of operating systems (Athey and Stern, 2014).

These findings suggest that no single factor explains the (lack of) use of open source software better than the lack of patenting, as an indicator of lack of technical sophistication within a country. Beyond that, all factors contribute to more proprietary (less open source) software, including higher income, less capable network (i.e., lower speeds), less capable inventors (i.e., fewer inventions), and a better environment for innovators. All these effects are moderate or imprecisely estimated, at best, and difficult to reconcile with one another.

Unlike the prior two estimates, these regressions cannot explain a high fraction of the variance in open source or proprietary web server software. This finding about open source shares, especially when set against the prior two findings about levels and per capita levels, suggests that we have not yet found the key drivers of open source web server market share.

V.4. Robustness

We tested the sensitivity of the estimates to a number of changes in the dataset and specification. The inferences largely hold, with one interesting exception that leads to an additional insight.

Table 9 tests the robustness of the findings to differences in economic development. Do high income countries differ from low income countries in their use of webservers? We split the sample in half, between countries with per capita income above and below the median.[30] The table presents new estimates of column two from Tables 6, 7, and 8. Column two is chosen due to its sample size, which involves 161 countries. The tests of the environment reduce the sample size too much to split the sample, so we do not test these variables. Even at a sample size of 80, however, this asks a lot of this data because the reduction in the size of the sample does effect the precision of the estimates.

Do the estimates (other than the constants) generally differ between high and low income countries? The estimates in Table 9 show that all the other coefficients do not statistically differ from each other, except in one case. The estimated coefficient for the log of price of broadband differs between below and above median income in the estimates for per capita servers and proprietary software share. In both cases, price matters for high income countries and not for low income countries.

[30] The split is at a medium income of $6350 per capita income, which is a “low-middle” income country.

Lower prices lead to more per capita servers and to a larger share of proprietary software. The elasticity estimate is large, at -0.80 for per capita server growth, and 0.62 for proprietary software shares. In both cases, these are large effects. It suggests lower broadband prices (expectedly) raise the number of per capita servers and (unexpectedly) raise the share of proprietary software. Lower prices also influence total servers to a similar extent in low and high income countries (see the previous two columns), so columns 1-4 together suggest a provocative finding. Lower broadband prices lead to more server growth, but only in high income countries does it lead to higher per capita servers. Also, in medium/high income countries growth leads to a larger share of proprietary web server software.

We tested for additional measures of network sophistication and encountered multicollinearity (results not shown here). We attempted to use IP addresses, hosts per country, internet users, or number of broadband subscriptions, as measures of network sophistication, but these highly correlate with the variables already in the regression.[31] No additional inference was possible. We also tested additional measures of the environment. The Competitiveness Global Institute has produced a range of indices for measuring the business environment (with sample sizes of 125). In addition to the reductions due to sample size, we encountered multicollinearity with the measure we use and could not make additional inferences in our data.

Table 10 summarizes the findings about the determinants of web servers. While the findings confirm most of the hypotheses, a number of notable hypotheses were rejected. The findings stress the importance of network quality and a country's technical skill. The estimates do not suggest much role for income in comparison to other factors. Particularly surprising, property rights and rule of law have little role on web server investment except in so far as it is shaped by the environment for innovation.

VI. Implications and Discussion

This study documented the size and dispersion of web servers across the globe, and analyzed the determinants. We analyzed over 24 million servers affiliated with the top three platforms. We document a pervasive spread of web servers to every country. Given the importance of the internet to

[31] Internet users and broadband subscriptions are available at World Bank Development Indicators. The IP addresses is available at (https://www.countryipblocks.net/allocation-of-ip-addresses-by-country.php), and hosts are available at (https://www.cia.gov/library/publications/the-world-factbook/rankorder/2184rank.html). These are very highly correlated with broadband speed and price, and, thus, contain no additional statistical information.

many economies, it is not surprising that such investment has spread. Yet, we also find that the United States continues to play an outsized role in delivering content, being home to 44% of global servers.

Open source web servers generate no revenue. Within the US, the lack of revenue does not correlate with lack of importance or relevance to the performance of the internet. This study reiterates a similar observation about web servers outside the US. This study shows that open source web servers comprise a large fraction of web server use in virtually every country with large investments in servers, its value reaching tens of billions of dollars around the globe.

We performed analysis to explain why webservers are unevenly distributed across the world. These findings have implications for policy to encourage internet content provided within a country's borders. These findings suggest that in lower income countries, it is not the income per se that holds back the growth in server software. While the technical sophistication of infrastructure does play a role, policies also must account for the technical sophistication of local inventors and the local labor market. Relatedly, these findings suggest low income countries may be dependent on content from high income countries, especially when low income countries have not invested in the other parts of ITC infrastructure.

Web servers are a big contributor to the IT economy around the globe, but remain invisible to traditional approaches for measuring GDP and the capital stock. Thus, our findings also have implications for policies related to economic accounting. Mismeasurement of open source can lead to misattribution in economic growth accounting. Standard economic measurement will attribute growth to the hardware and not properly attribute the gains to the unmeasured input, and it may overlook the essential role of sophisticated human inputs. However, our findings also point toward a silver lining. The presence of web servers correlates with higher quality networks and the technical sophistication of the local population. That finding carries with it the potential to proxy for missing web software with statistical methods. That is an important open question and worthy of further research.

These findings also inform policies for encouraging local web content. Policy should not expect local content to arise as a byproduct of income growth. Instead, policy should focus on growing a more sophisticated network and increasing the sophistication of local technical talent. Our findings also suggest that developed countries could encourage more local content through lower broadband prices, but lower income countries cannot. We do not find much support for policies that encourage local content through enforcing property rights.

The study motivates several open questions. For example, we find that web content and pirated content are not influenced by the same environmental factors. That is surprising and raises questions about the additional role of labor market sophistication. Additionally, while there is a large literature on the production of open source software, this study shows that use correlates with the quality of related complementary networks. If those factors explain difference across countries, then what determines the geographic dispersion of its deployment within a country, for example, from one city or region to another? Additionally, this study examined one cross section, and not any upgrading over time. That motivates questions about how to weight different vintages of software and compare their upgrading patterns over time and across regions. The methods for identifying location would need modification for such a purpose and that remains an open topic. Finally, this study examined but one piece of open source software and it is widely regarded as *less* popular than Linux. We expect Linux, as well as other open source software, to amount to large values of otherwise invisible value. However, the precise amounts awaits further methods to enable census of their use and deployment.

References

Ackermann, Klaus, and Simon D. Angus. "A resource efficient big data analysis method for the social sciences: the case of global IP activity." *Procedia Computer Science* 29 (2014): 2360-2369.

Ackerman, Klaus, Simon D Angus, and Paul A Raschky, 2017. The Internet as Quantitative Social Science Platform: Insights from a Trillion Observations, ArcX.

Aguiar, Luis, and Joel Waldfogel, 2014. "Digitization, Copyright, and the Welfare Effects of Music Trade," SSRN: https://ssrn.com/abstract=2603238 or http://dx.doi.org/10.2139/ssrn.2603238

Athey, Susan, and Scott Stern. 2015. The Nature and Incidence of Software Piracy: Evidence from Windows. Avi Goldfarb, Shane Greenstein, and Catherine Tucker, Editors, *Economic Analysis of the Digital Economy,* NBER conference book, University of Chicago Press.

Blum, Bernardo, and Avi Goldfarb. 2006. "Does the internet defy the law of gravity?" *Journal of International Economics*, 70(2), 384-405.

Botnet, Carna. "Internet Census 2012: Port scanning/0 using insecure embedded devices." *URL http:// census2012.sourceforge.net/paper.html* (2013).

Delgado, Mercedes, Christian Ketels, Michael E Porter, and Scott Stern, 2012. The Determinants of National Competitiveness. National Bureau of Economic Research, Working Paper # 18249.

DiBona, Chris, Sam Ockman, Mark Stone, 1999. *Voices from the Revolution*, O'Reilly Media; Sebastopol, CA.

Fershtman, Chiam, and Neil, Gandal, 2011, "A Brief Survey of the Economics of Open Source Software," in (eds) L Blume, and S Durlauf, *The new Palgrave Dictionary of Economics.* Palgrave Macmillan.

Gharaibeh, Manaf, et al., 2017. "A look at router geolocation in public and commercial databases." *Proceedings of the 2017 Internet Measurement Conference*. ACM, 2017.

Gomez-Herrera, Estrella, and Bertin Martins, 2014, "The Driver and Impendiments from Cross Border E-Commerce in the EU," *Information Economics and Policy,* V28, pp. 83-96.

Greenstein, Shane, 2015. *How the Internet Became Commercial: Innovation, Privatization, and the Birth of a new Network,* Princeton University Press.

Greenstein, Shane, and Frank Nagle, 2014. "Digital Dark Matter and the Economic Contribution of Apache," *Research Policy*. v43. Pp 623-631.

King, Gary, Jennifer Pan, and Margaret E. Roberts, 2014. "Reverse-engineering censorship in China: Randomized experimentation and participant observation." *Science* 345.6199: 1251722.

Lakhani, K., von Hippel, E., 2003. How open source software works: free user-to-user assistance. Research Policy 32 (6), 923–943.

Lerner, Josh, and Mark Schankerman, 2010. *The Comingled Code: Open Source and Economic Development.* MIT Press.

Murciano Goroff, Raviv, 2018. Missing Women in Tech: The Labor Market for Highly Skilled Software Engineers, working paper, standard.edu/~ravivmg/

Nagle, Frank, 2018. "Learning By Contributing: Gaining Competitive Advantage Through Contribution to Crowdsourced Public Goods." *Organization Science*.

Nagle, Frank, 2017. "Open Source Software and Firm Productivity," *Management Science.*

Nordhaus, William D., 2006. Principles of national accounting for nonmarket accounts. In:Jorgenson, D.W., Landefeld, J.S., Nordhaus, W.D. (Eds.), *A New Architecture for the US National Accounts.* University of Chicago Press, Chicago, IL.

OECD (2013). *Communications Outlook*, OECD Publishing, Paris. https://doi.org/10.1787/comms_outlook-2013-en.

Papke, Leslie E., and Jeffrey M. Woolridge. 1996. Econometric methods for fractional response variables with an application to 401(k) plan participation rates, *Journal of Applied Econometrics.* 11(6). Pp. 619-632.

Wen Wen, Marco Ceccagnoli, Chris Forman, 2016, "Opening up Intellectual Property Strategy: Implications for Open Source Entry by Start-up Firms." *Management Science. 62(9),* pp. 2668-2691.

Table 1. **US and next top 30 suppliers of global webservers**

Country	Apache count	Percent Apache	IIS count	Percent of IIS	Nginx count	Percent Ngnix	IIS Country Share	$B Open Source Value
Total	15401242		6924722		1956125		26.2%	$42.11
USA	6832207	44.4%	3212926	46.4%	540841	27.6%	30.4%	$18.36
China	565167	3.7%	913792	13.2%	507103	25.9%	46.0%	$3.44
Germany	1199952	7.8%	238369	3.4%	78299	4.0%	15.7%	$2.63
Japan	912793	5.9%	99872	1.4%	20242	1.0%	9.7%	$1.79
UK	504847	3.3%	329551	4.8%	35583	1.8%	37.9%	$1.51
Netherlands	480732	3.1%	127774	1.8%	72031	3.7%	18.8%	$1.18
Canada	347669	2.3%	228870	3.3%	14806	0.8%	38.7%	$1.03
France	485455	3.2%	90083	1.3%	172069	8.8%	12.0%	$1.30
Russian Fed	344019	2.2%	64618	0.9%	164541	8.4%	11.3%	$0.99
Brazil	239268	1.6%	84632	1.2%	6734	0.3%	25.6%	$0.57
Korea, Rep.	183095	1.2%	133155	1.9%	5706	0.3%	41.4%	$0.56
Australia	164454	1.1%	144161	2.1%	47609	2.4%	40.5%	$0.62
Italy	173034	1.1%	98238	1.4%	3694	0.2%	35.7%	$0.48
Taiwan	157068	1.0%	86843	1.3%	6860	0.4%	34.6%	$0.43
Poland	216590	1.4%	20534	0.3%	18250	0.9%	8.0%	$0.44
Sweden	167316	1.1%	57867	0.8%	8411	0.4%	24.8%	$0.41
Spain	149469	1.0%	49669	0.7%	5076	0.3%	24.3%	$0.35
Honk Kong	95415	0.6%	76207	1.1%	8503	0.4%	42.3%	$0.31
India	87953	0.6%	71641	1.0%	18031	0.9%	40.3%	$0.31
Israel	115579	0.8%	30423	0.4%	1314	0.1%	20.7%	$0.26
Turkey	69408	0.5%	64088	0.9%	4095	0.2%	46.6%	$0.24
Switzerland	84846	0.6%	35588	0.5%	2479	0.1%	29.0%	$0.21
Czech Rep	96815	0.6%	19669	0.3%	15043	0.8%	15.0%	$0.23
Ukraine	102191	0.7%	12392	0.2%	28174	1.4%	8.7%	$0.25
Thailand	74160	0.5%	32164	0.5%	2052	0.1%	29.7%	$0.19
Ireland	74023	0.5%	24360	0.4%	15580	0.8%	21.4%	$0.20
Denmark	48656	0.3%	42210	0.6%	1673	0.1%	45.6%	$0.16
South Africa	31992	0.2%	35750	0.5%	1042	0.1%	52.0%	$0.12
Bulgaria	36363	0.2%	5880	0.1%	93391	4.8%	4.3%	$0.24
Latvia	11566	0.1%	2603	0.0%	7112	0.4%	12.2%	$0.04

Table 2a. List of countries included in sample

Afghanistan, Albania, Algeria, Andorra, Angola, Antigua and Barbuda, Argentina, Armenia, Aruba, Australia, Austria, Azerbaijan, Bahamas, Bahrain, Bangladesh, Barbados, Belarus, Belgium, Belize, Benin, Bermuda, Bhutan, Bolivia, Bosnia and Herzegovina, Botswana, Brazil, British Virgin Islands, Brunei Darussalam, Bulgaria, Burkina Faso, Burundi, Cabo Verde, Cambodia, Cameroon, Canada, Cayman Islands, Central African Republic, Chad, Chile, China, Colombia, Comoros, Congo Dem. Rep., Congo, Rep., Costa Rica, Cote d'Ivoire, Croatia, Cuba, Curacao, Cyprus, Czech Republic, Denmark, Djibouti, Dominica, Dominican Republic, Ecuador, Egypt, Arab Rep., El Salvador, Equatorial Guinea, Eritrea, Estonia, Ethiopia, Faroe Islands, Fiji, Finland, France, French Polynesia, Gabon, Gambia, Georgia, Germany, Ghana, Gibraltar, Greece, Greenland, Grenada, Guam, Guatemala, Guinea, Guinea-Bissau, Guyana, Haiti, Honduras, Hong Kong, Hungary, Iceland, India, Indonesia, Iran, Iraq, Ireland, Isle of Man, Israel, Italy, Jamaica, Japan, Jordan, Kazakhstan, Kenya, Kiribati, Democratic People's Republic of Korea, Republic of Korea, Kuwait, Kyrgyz Republic, Lao PDR, Latvia, Lebanon, Lesotho, Liberia, Libya, Liechtenstein, Lithuania, Luxembourg, Macao, Macedonia FYR, Madagascar, Malawi, Malaysia, Maldives, Mali, Malta, Marshall Islands, Mauritania, Mauritius, Mexico, Micronesia, Moldova, Monaco, Mongolia, Montenegro, Morocco, Mozambique, Myanmar, Namibia, Nauru, Nepal, Netherlands, New Caledonia, New Zealand, Nicaragua, Niger, Nigeria, Northern Mariana Islands, Norway, Oman, Pakistan, Palau, Panama, Papua New Guinea, Paraguay, Peru, Philippines, Poland, Portugal, Qatar, Russian Federation, Rwanda, Samoa, San Marino, Sao Tome and Principe, Saudi Arabia, Senegal, Serbia, Seychelles, Sierra Leone, Singapore, Sint Maarten (The Dutch Part), Slovak Republic, Slovenia, Solomon Islands, Somalia, South Africa, South Sudan, Spain, Sri Lanka, St. Kitts and Nevis, St. Lucia, St. Martin (French part), St. Vincent and the Grenadines, Sudan, Suriname, Swaziland, Sweden, Switzerland, Syrian Arab Republic, Tajikistan, Tanzania, Taiwan, Thailand, Timor-Leste, Togo, Tonga, Trinidad and Tobago, Tunisia, Turkey, Turkmenistan, Turks and Caicos Islands, Tuvalu, Uganda, Ukraine, United Arab Emirates, United Kingdom, United States, Uruguay, US Minor Outlying Islands, Uzbekistan, Vanuatu, Venezuela, Vietnam, West Bank and Gaza, Yemen Republic, Zambia, Zimbabwe.

Table 2b. Statistics for sample

Variable	Number of countries	Mean	s.d.	Min	Max
Number of Microsoft servers	213	32402.22	230225.8	3	3212926
Number of Apache servers	213	70851.62	483434.0	3	6832207
Number of Nginx servers	213	9178.21	53860.4	0	540841
Total number of servers in country	213	112432.1	752561.3	6	10600000
Ln of all servers	213	7.79	2.87	1.79	16.18
Ln of all servers per capita	210	-7.37	2.56	-15.23	-2.15

Table 3. Summary of hypotheses

Determinant	**Measure**	**Hypotheses: Level of server software.**	**Hypothesis: Levels of per capita sw.**	**Hypotheses: Share of proprietary sw.**
Economic development	Ln of Per capita income	Positive	Positive	Positive
	Perc of urban electricity	Positive	Positive	Positive
Network quality	Ln of broadband speed	Positive	Positive	No hypothesis
	Ln of broadband price	Negative	Negative	No hypothesis
Technical sophistication	Patents per capita	Positive	Positive	Negative
	Dummy for any patents	Negative	Negative	Positive
Institutional environment	Index of Property Rights	Positive	Positive	Positive
	Index of Innovation	Positive	Positive	Positive
	Rule of Law Index	Positive	Positive	Positive

Table 4. Descriptive statistics.

Variable	**Number of obs**	**Mean**	**s.d.**	**Min**	**Max**
Economic development					
Ln of GDP per capita	192	8.68	1.50	5.50	11.90
% urban pop w/ electricity	205	89.31	20.59	8.98	100.00
Network quality					
Ln of broadband speed	187	0.12	1.36	-1.39	4.61
Ln of broadband price	176	3.29	0.87	-0.04	7.38
Technical sophistication					
Ln of patents per capita	213	-8.52	6.16	-18.07	0
No patents	213	0.30	0.46	0.00	1.00
Institutional environment					
Heritage prop rights index	177	59.59	11.54	1.00	89.90
GC Index of innovation	138	-0.01	0.29	-0.67	0.68
Rule of Law index	112	0.57	0.14	0.28	0.89

Table 5

Endogenous variable: Ln (servers)

Economic development					
Ln of GDP per capita	**0.75**	-0.21	0.01	-0.22	0.28
	0.14	*0.21*	*0.23*	*0.25*	*0.30*
% urban pop w/electricity	**0.04**	0.01	0.00	0.01	0.01
	0.01	*0.01*	*0.01*	*0.01*	*0.02*
Network quality					
Ln of broadband speed		**0.41**	**0.35**	**0.29**	0.27
		0.13	*0.13*	*0.14*	*0.15*
Ln of broadband price		**-0.68**	**-0.70**	**-1.01**	**-0.54**
		0.20	*0.21*	*0.27*	*0.28*
Technical sophistication					
Ln of patents per capita		**0.50**	**0.49**	**0.41**	**0.64**
		0.11	*0.11*	*0.13*	*0.15*
No patents		**-9.43**	**-9.31**	**-7.43**	**-11.53**
		1.67	*1.70*	*2.13*	*2.18*
Institutional environment					
Heritage prop rights index			-0.01		
			0.02		
GC Index of innovation				**2.26**	
				1.00	
Rule of Law index					**-5.56**
					2.21
Constant	**-1.82**	**18.72**	**18.12**	**18.86**	**19.13**
	0.97	*3.02*	*3.20*	*3.52*	*4.15*
Number of countries	189	161	153	125	104
R-Squared	0.38	0.64	0.67	0.64	0.66

Bold means significance at 5% level.

Standard errors in italics

Table 6

Endogenous variable: Ln (Servers per capita)

Economic development					
Ln of GDP per capita	**1.16**	**0.63**	**0.55**	**0.63**	**0.84**
	0.08	*0.13*	*0.15*	*0.14*	*0.20*
% urban pop w/electricity	**0.03**	**0.03**	**0.02**	0.01	**0.03**
	0.01	*0.01*	*0.01*	*0.01*	*0.01*
Network quality					
Ln of broadband speed		**0.18**	0.14	**0.25**	0.19
		0.08	*0.08*	*0.08*	*0.10*
Ln of broadband price		-0.11	-0.16	**-0.50**	-0.11
		0.12	*0.13*	*0.16*	*0.18*
Technical sophistication					
Ln of patents per capita		**0.31**	**0.32**	**0.36**	0.17
		0.07	*0.07*	*0.08*	*0.10*
No patents		**-4.50**	**-4.58**	**-5.54**	**-2.95**
		1.03	*1.08*	*1.22*	*1.42*
Institutional environment					
Heritage prop rights index			**0.02**		
			0.01		
GC Index of innovation				-0.25	
				0.57	
Rule of Law index					0.45
					1.44
Constant	**-19.78**	**-10.78**	**-10.87**	**-7.50**	**-14.79**
	0.54	*1.88*	*2.02*	*2.02*	*2.70*
Number of countries	189	161	153	125	104
R-Squared	0.74	0.80	0.81	0.85	0.81

Bold means significance at 5% level.

Standard errors in italics

Table 7

Descriptive stats

Variable	Number of countries	Mean	s.d.	Min	Max
Share of servers with Microsoft software	213	0.35	0.15	0.01	0.90
Ln (Microsoft share) – ln[1 - Microsoft share)	213	-0.76	0.90	-5.15	2.19

Table 8

Endogenous variable: Ln [Microsoft share/(1 – Microsoft share)]

Economic development					
Ln of GDP per capita	0.10	**0.35**	**0.26**	0.15	0.32
	0.06	*0.11*	*0.12*	*0.13*	*0.17*
% urban pop w/electricity	-0.01	-0.01	0.00	0.01	-0.01
	0.00	*0.01*	*0.01*	*0.01*	*0.01*
Network quality					
Ln of broadband speed		-0.08	-0.07	-0.12	-0.04
		0.07	*0.07*	*0.07*	*0.08*
Ln of broadband price		0.05	0.07	0.23	0.00
		0.10	*0.11*	*0.14*	*0.15*
Technical sophistication					
Ln of patents per capita		**-0.13**	**-0.12**	**-0.14**	-0.09
		0.06	*0.06*	*0.07*	*0.08*
No patents		**2.18**	**2.03**	**2.24**	1.83
		0.85	*0.91*	*1.09*	*1.21*
Institutional environment					
Heritage prop rights index			0.01		
			0.01		
GC Index of innovation				0.62	
				0.51	
Rule of Law index					-0.44
					1.23
Constant	**-1.05**	**-5.15**	**-5.12**	**-5.40**	-3.83
	0.37	*1.54*	*1.70*	*1.80*	*2.31*
Number of countries	189	161	153	125	104
R-Squared	0.02	0.09	0.08	0.12	0.08

Standard errors in italics

Table 9. Robustness tests

	Ln (servers)		Ln (Servers per capita)		Ln [Microsoft share/(1 – Microsoft share)]	
	Below median per cap income	**Above median per cap income**	**Below median per cap income**	**Above median per cap income**	**Below median per cap income**	**Above median per cap income**
Income						
Ln of GDP per capita	0.10	-0.85	**1.17**	0.22	0.20	**0.51**
	0.31	*0.49*	*0.24*	*0.20*	*0.19*	*0.19*
% Urban pop w/electricity	0.00	0.09	0.01	-0.02	0.00	-0.01
	0.01	*0.06*	*0.01*	*0.02*	*0.01*	*0.02*
Network quality						
Ln of broadband speed	0.29	0.39	0.10	**0.20**	0.03	-0.11
	0.18	*0.21*	*0.14*	*0.08*	*0.11*	*0.08*
Ln of broadband price	**-0.71**	-0.51	0.01	**-0.80**	-0.03	**0.62**
	0.19	*0.60*	*0.15*	*0.24*	*0.12*	*0.24*
Technical sophistication						
Ln of patents per capita	**0.30**	**0.68**	**0.37**	**0.51**	-0.06	**-0.25**
	0.15	*0.21*	*0.11*	*0.08*	*0.09*	*0.08*
No patents	**-6.37**	**-12.25**	**-5.58**	**-5.98**	1.34	**3.09**
	2.23	*2.76*	*1.77*	*1.12*	*1.41*	*1.09*
Constant	**13.86**	**18.86**	**-13.28**	1.85	-2.90	**-9.23**
	3.63	*8.15*	*2.88*	*3.31*	*0.05*	*0.31*
Number of countries	81	80	81	80	81	80
R-Squared	0.61	0.47	0.65	0.72	0.05	0.31

Bold means significance at 5% level.

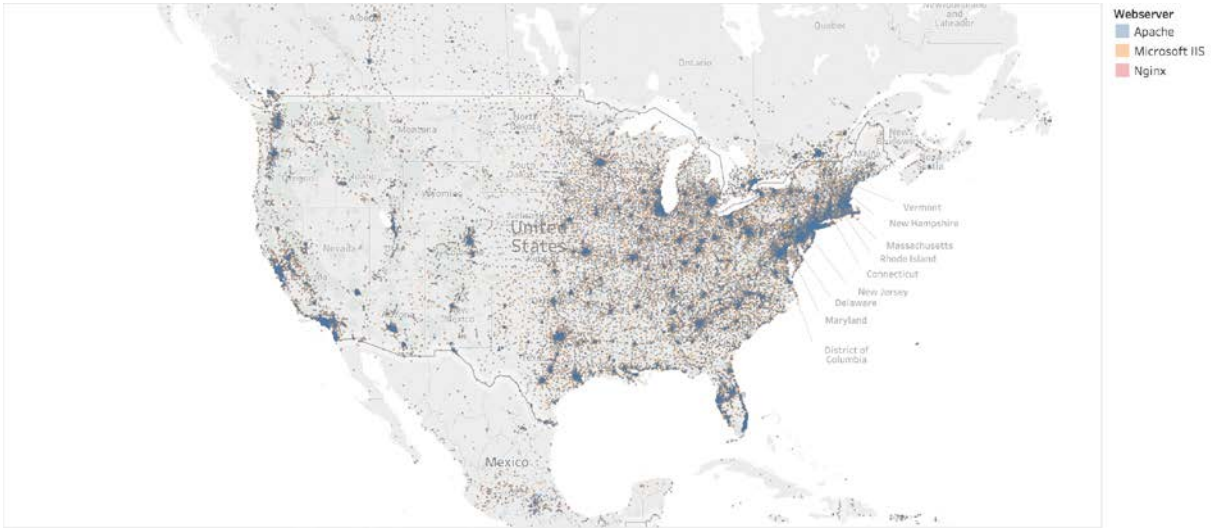
Standard errors in italics

Table 10. Summary of results

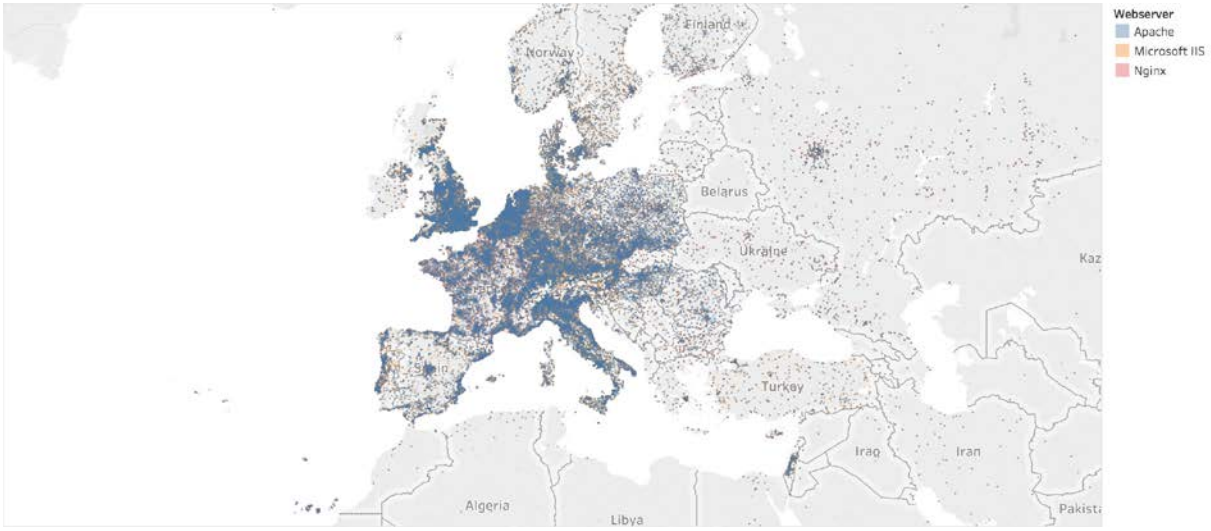
Determinant	**Proxy**	**Hypotheses: Level of server software.**	**Hypothesis: Levels of per capita sw.**	**Hypotheses: Share of proprietary sw.**
Economic development	Ln of Per capita income	Pos (reject)	Pos (accept)	Pos (reject)
	Perc of urban electricity	Pos (reject)	Pos (weak accept)	Pos (reject)
Network quality	Ln of broadband speed	Pos (accept)	Pos (reject)	None
	Ln of broadband price	Neg (accept)	Neg (accept for high income)	None (Neg for high income)
Technical sophistication	Patents per capita	Pos (accept)	Pos (accept)	Neg (accept)
	Dummy for any patents	Neg (accept)	Neg (accept)	Pos (accept)
Institutional environment	Index of Property Rights	Pos (reject)	Pos (accept)	Pos (reject)
	Index of Innovation	Pos (accept)	Pos (reject)	Pos (reject)
	Rule of Law Index	Pos (reject)	Pos (reject)	Pos (reject)

Figure 1. Maps

USA:



Europe:



World:

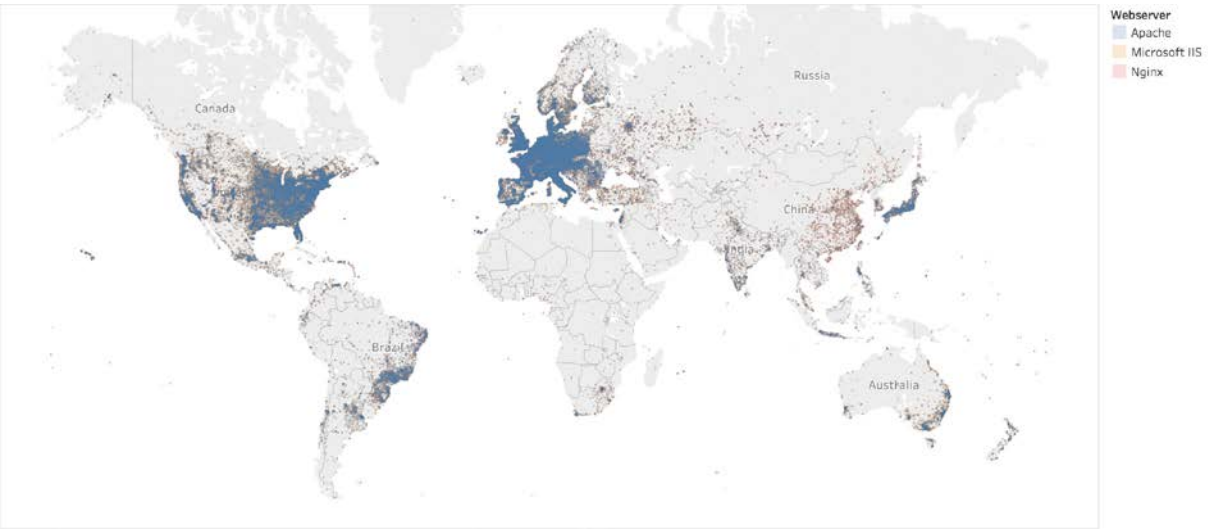
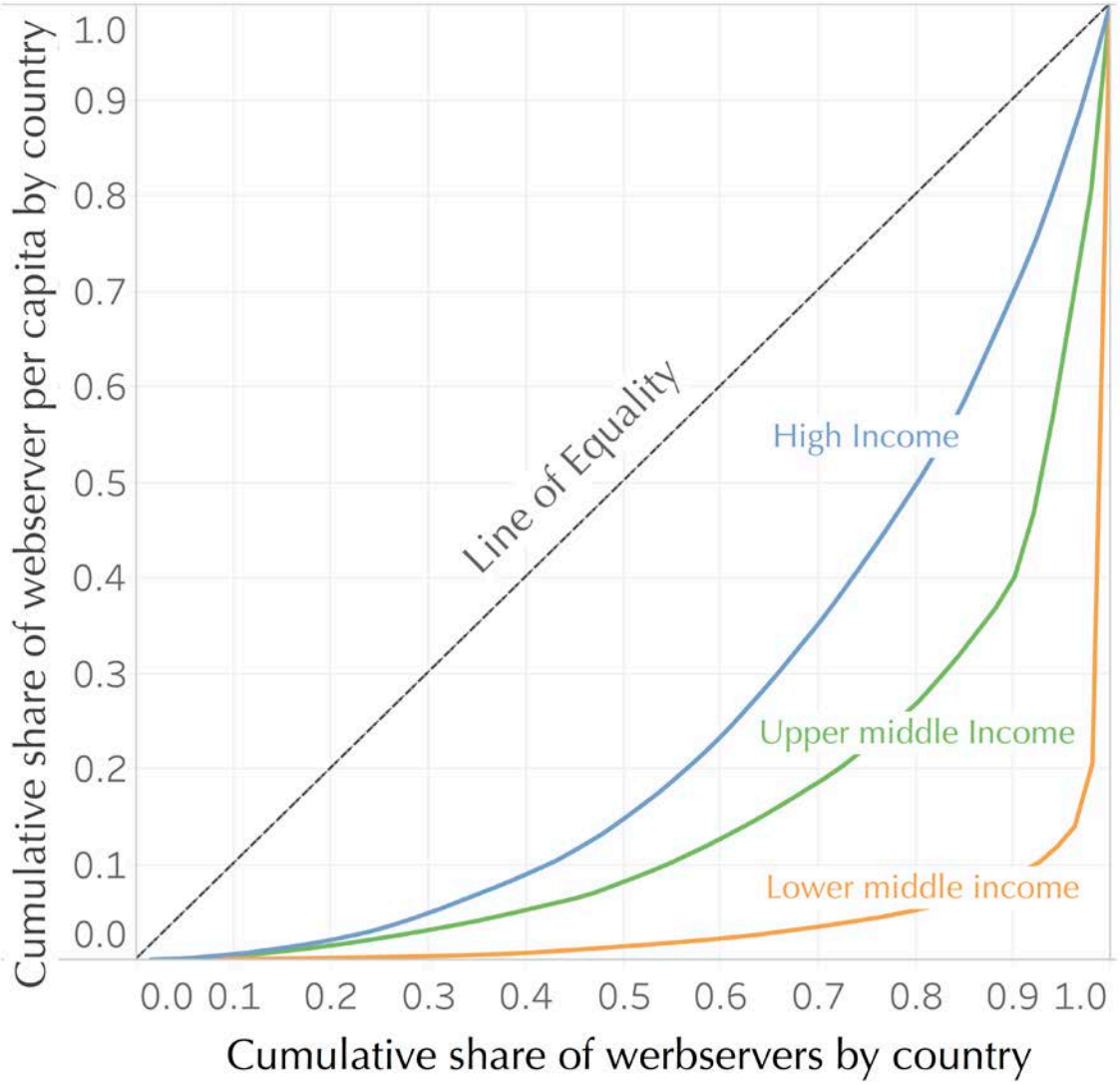

Webserver
Apache
Microsoft IIS
Nginx
Canada
Russia
China
India
Brazil
Australia

Figure 2:


Cumulative share of webserver per capita by country
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Line of Equality
High Income
Upper middle Income
Lower middle income
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Cumulative share of werbservers by country