

Department of Land Economy

Environment, Law & Economics



Working Paper Series

No. 2016-04

Title: Monocentric Cyberspace: The Primary Market for Internet Domain Names

Authors: Thies Lindenthal

Affiliation: University of Cambridge, Department of Land Economy

Contact corresponding author: Thies Lindenthal, htl24@cam.ac.uk

Monocentric Cyberspace: The Primary Market for Internet Domain Names

Thies Lindenthal, Department of Land Economy, University of Cambridge, UK. E-mail: htl24@cam.ac.uk

Abstract:

Cyberspace is no different from traditional cities, at least in economic terms. Urban economics governs the creation of new space on the Internet and explains location choices and price gradients in virtual space. This study explores registration dynamics in the largest primary market for virtual space: Internet domain names. After developing a framework for domain registrations, it empirically tests whether domain registrations are constrained by the depletion of unregistered high quality domain names. Estimations based on registrations of COM domain names suggest that the number of domains expands substantially slower than the growth in overall demand for domain space. Supplying alternative domain extensions can relax the shortage in domains in the short term.

Keywords: Virtual Land-Rush Markets, Primary Market for Cyberspace, Registrations of Internet Domain Names, Supply Constraints

Acknowledgements:

My postdoctoral research project received substantial financial contributions by the Internet Corporation of Assigned Names and Numbers (ICANN). I would like to explicitly thank Cyrus Namazi and Mike Zupke for supporting this research project and for giving me the opportunity to discuss preliminary results with ICANN. Albert Saiz, David Gelter, Piet Eichholtz, and Liao Wen-Chi are thanked for their insightful feedback on earlier versions of this paper.

One of the defining myths of the World Wide Web is that ever-increasing levels of electronic communication and interconnectedness will ultimately overcome physical distance. Higher and higher degrees of virtual omnipresence hypothetically raze traditional boundaries, reshape the spaces we live and work in (among others, Gaspar and Glaeser 1998; Ihlanfeldt 1995; Pascal 1987; Tranos and Nijkamp 2013) and will eventually culminate in the “death of distance” (Cairncross 1997). Put more provocatively, the world is said to “flatten” (Friedman 2005),

Putting aside all information technology exuberance, location will continue to matter. Even if we relocated all our work, social interactions, shopping, and leisure activities into cyberspace¹ and left our physical shells in self-sufficient containers that provide the same amenities anywhere – our world would still be governed by distances and locations, albeit in new dimensions.

One of the largest markets for locations in spaces beyond the traditional coordinate systems, is the market for Internet domain name registrations. Domain names map natural language character strings, which can be easily memorized by humans, to technical network addresses on the Internet, which tend to be hard to recall². Strictly speaking, they are just pointers to data and online services, but effectively they provide locations that enable humans to navigate the web (Mueller 1998).

So far, the primary market for domain names is uncharted territory in academic literature. This paper is the first to explore the determinants of domain registrations using an adaptation of the bedrock of Urban Economics, the classic monocentric city model (Alonso 1964; Mills 1972; Muth 1969). In this model, a city is located in a featureless plain and all employment is located in the central business district (CBD), to which all residents commute regularly. When selecting a lo-

¹ Batty's (1997) virtual geography terminology distinguishes between *cyberplaces*, in which the built form and electronic networks are interdependent, and *cyberspace*, which is fully detached from physical space.

²For a primer on domain names see Mueller (1998) or Moss and Townsend (1997).

cation to live, residents trade off the cost of commuting as a function of the distance to the CBD, vis-a-vis the consumption of other goods. In equilibrium, rent level rise as the distance to CBD falls.

In virtual space, a comparable rent gradient can be observed: Sought-after virtual “downtown locations” transact for several million USD (The Economist Online 2010), similar to locations in the centers of brick-and-mortar cities, while domains located in the cyber-periphery trade for a fraction of the central spaces’ values only (Lindenthal 2014). Analogous to the monocentric city model, differences in prices of virtual locations can be explained by the differences in distance to a central location. In virtual space, however, all users reside in the same location and travel from this universal origin to different virtual destinations to access information or services. While it takes only a few keystrokes to direct a browser to a new location, the associated effort varies between different destinations. The cost of cyber-commuting depends on the linguistic attributes of the destinations which differ in familiarity, pronounceability, and memorability of the domain names (for an excellent review of proper name memory see Cohen and Burke 1993). Neuroscience studies have shown that the human memory stores and retrieves different types of concepts, words and names not only through a variety of neural systems but also in separate locations within the brain (e.g. Binder et al. 2009; Damasio et al. 2004; Humphreys and Forde 2001). Cyberspace may have overcome traditional concepts of physical distances, but not all locations are equally easy to communicate, memorize, recall, and type. This paper develops measures of some of these distances and empirically show that word-specific cyber-commuting costs determine domain registrations. Documenting virtual distances and an equivalent to commuting costs is a prerequisite when applying the monocentric city model to virtual space.

At first sight, the seemingly unlimited³ number of domains that still can be registered against a low and exogenously determined registration fee appears to be at odds with the monocentric city model where land in the CBD is naturally scarce: If supply of space was unconstrained every website could have the most desirable location and should be valued at the same nominal registration fee. Again, showing that supply of space is constrained is another necessary step towards a monocentric virtual city model.

Three stylized facts suggest that domain names actually are indeed a limited resource. First, the annual increase in the number of registered domain names has been slowing down in the last years while the global Internet user base has been growing at an increasing pace. Figure 1 shows that up to 2007, the universe of COM domain names has expanded at a similar speed to the number of world residents with Internet access. From 2007 onwards, the annual additions to the domain stock lag the overall trend which could be caused by fewer and fewer high quality keywords remaining available for registration. With more than 271 million registered domains at the end of 2013 (Verisign 2014a), it is not easy to find a good domain that has not been claimed yet.

===== *Insert Figure 1 about here* =====

Second, the re-sale prices of registered domain names have risen 63% from 2006 through 2012 (Lindenthal 2014). These rising prices are indicative of the demand for virtual locations outpacing the supply of available attractive names and that competition drives up prices for “central” domains.

³The Internet standard RFC 1035 (<http://tools.ietf.org/html/rfc1035>) and its subsequent updates specify the rules for domain name supply. Each *label* a domain name consists of (labels are separated by dots) can contain up to 63 octets, allowing for e. g. 26⁶³ combinations of the latin characters A to Z in the Second Level Domain. Adding numbers, dashes and millions of internationalized domain names further increases this already massive number of potential domain names.

Third, the Internet Corporation for Assigned Names and Numbers (ICANN), the non-profit organization regulating the Domain Name System (DNS), has initiated the release of 1400 new Top Level Domains (TLD) that will augment the current selection of 22 global extensions like COM, EDU, NET or ORG (Internet Corporation for Assigned Names and Numbers 2014). Entrepreneurs have made large bets on the right to monetize these new swathes of cyberspace, putting down about 350 million USD in application fees (Internet Corporation for Assigned Names and Numbers 2013) and investing billions more for the necessary infrastructure needed to manage the new space. Clearly, they have trust in being able to serve a previously unmet demand.

Estimating the extent of supply constraints is not only interesting for urban economists, but also a timely and relevant challenge for policy makers pondering whether the current domain name system serves the Internet optimally and for the business community trying to serve any unmet demand for space.

The next section of this paper suggests a framework that can empirically test supply constraints, price gradients and demand curves for Internet domain names. Subsequently, data on domain registrations will be introduced before the last two sections present the results of the empirical estimations and a conclusion.

THE DETERMINANTS OF DOMAIN NAME REGISTRATIONS

Three factors determine the total number of registered domains: First, the ultimate driver of domain name registrations is the demand for virtual space by businesses, organizations or individuals that offer Internet-based information and services to Internet end users. More than 20 years after the inception of the Internet, the total head count of these virtual dwellers (*Pop*) is still expanding rapidly. While the exact amount of space demanded per dweller is difficult to quantify and also

likely to change over time, it is safe to assume that Pop and total demand for virtual space are positively correlated: For instance, doubling the number of virtual dwellers is expected to lead to twice the demand for space (keeping all other factors equal).

Second, the registration fees and other fixed costs (K_{reg}) associated with owning a domain name are negatively related to total registration numbers. The fixed costs are comprised of the wholesale domain registration fees charged by the company that administers the domain registry, the markup added by competing middlemen re-selling domains to end users, and by additional costs for hosting and related services. While the direct costs of owning and hosting a domain have been falling year after year due to intense competition between service providers⁴, it is safe to assume that K is identical for all registrations in a cross-sectional study.⁵

Ultimately, each end user of a website needs exercise an effort E to access an online location. This commuting cost depends on domain specific factors like the recognizability and ease of recollection of a specific name and also on the general competitive position of domains versus other forms of virtual space. If the required effort of commuting to a location is low, owning this location is desirable as it is possible to attract end users easily. Locations with high required efforts are less attractive since fewer end users visit. New domains get registered as long as the utility gained from owning marginal domains given a marginal effort level $E_{marginal}$ required by any user exceeds registration costs. In sum, the total number of registrations of domain names can be formalized as

$$Registrations = a \frac{bPop}{(cK)(dE_{marginal})} \quad (1)$$

⁴Anecdotal evidence on long term trends in web hosting is presented by *Royal Pingdom* <http://royal.pingdom.com/2008/02/19/web-hosting-now-vs-10-years-ago/>

⁵Registration fees should not be confused with resale values in secondary markets. At any given time, wholesale registration fees are similar and do not depend on the inherent quality of the domain. Resale prices in the secondary market, however, vary greatly and the variation in values can be partially explained in hedonic regressions (Lindenthal and Loebbecke 2014) or a repeat sales analysis (Lindenthal 2014).

with a , b , c , d being scaling parameters accounting for the overall attractiveness of domain names versus other forms of virtual space (a) and the elasticities of registrations with respect to general demand (b), registration costs (c), and marginal efforts or commuting costs (d).

The marginal effort $E_{marginal}$ is assumed to increase in registration numbers. Domain name registrations exhibit a pecking order regarding domain quality: names that had been registered relatively early tend to be of higher quality than those registered later. Marginal domain registrations, on average, contain out of more characters, are less descriptive and more difficult to memorize than the existing stock, requiring higher efforts by end users as registrations increase. Those high quality and easy to access locations that are claimed first in land rush markets tend to trade for higher values in secondary markets subsequently.

The marginal level of effort required by users as more and more lower quality domains get registered, can be generalized as

$$E_{marginal} = g \text{ Registrations}^h, \quad (2)$$

where g and h are scaling factors. The choice of a power function is motivated by Zipf's observation that the frequency at which a word is used is inversely proportional to this word's rank in the frequency table (Zipf 1936) and that the rank is a particular power function of word frequency (Zipf 1949). If domains are registered along the rank suggested by the keyword frequency table, the marketing potential of domains will also follow a power law. This notion is supported by Cunha et al. (2011) finding that the frequencies of Twitter hashtags are governed by a Zipfian power distribution.

Assuming identical registration costs for all domains, and plugging (2) into (1) can be solved for Registrations and simplified to

$$\text{Registrations} = m\text{Pop}^n, \quad (3)$$

where where m and n are products of earlier used constants (and therefore constants as well): $m = (ab/cKg)^{-(1+h)}$ and $n = -(1+h)$.

This study assumes the same level of use intensity for all domains. While the classical, traditional monocentric model does include variable density, that is not a necessary feature of the monocentric model. All the essential elements of the monocentric model still come through with a fixed, constant density as shown in Geltner et al. (2001, Chapter 4). In addition, Lindenthal and Loebbecke (2014) have already documented that more valuable domains are more likely to be developed into more extensive websites, which represents a higher use intensity or “density” compared to registrations of lower quality domains without further development.

Owner-operated websites are not the only form of cyberspace available to virtual dwellers. Alternatively, they can connect with their audiences through shared spaces like social media platforms, wikis, online market places or direct communications and marketing. For instance, the increasing role of social networks in connecting companies and its customers could weaken the demand for domains in general. A local business might find it more cost-effective to promote its Facebook profile instead of steering customers towards their website. Reversed, changes to search engine algorithms could make it easier for users to find relevant content on millions of individual websites, tilting the balance in favor of owning domains. The competitive position of domains versus other options is, among other factors, accounted for in factor m .

This framework also allows investigating demand levels for segments of domains by employing subset specific values for E . For instance, the relative commuting costs for a domain under the

COM Top Level Domain (TLD) could differ from the cost of accessing a NET or ORG domain, resulting in the different demand levels for each TLD, documented by Lindenthal (2014).

The empirical part of the paper splits a cross-section of domain registrations into subsets for which the level of demand Pop is quantifiable and the number of registered domains is known. The fixed cost K is identically distributed for all domains in cross-section and can therefore be omitted. In a future study, the price sensitivity of domain registrations could be estimated by analyzing longitudinally different values for K .

Figure 2 visualizes the approach: For each group of domains, the intersection of the demand curve $D1$ and $D2$ and supply curve $S1$ can be observed as the number of registered domains, $R1$ and $R2$. Demand for domains from group 2 is higher than demand for domains from group 1, as indicated by an upward shifted demand curve and higher values for $R2$ than for $R1$. This analysis is only feasible if data on fundamental demand levels and registration numbers per segment can be directly observed.

===== *Include Figure 2 about here* =====

The relationship between the number of potential domain registrants Pop and actual *Registrations* can now be estimated empirically in the following log-log regression specification:

$$\ln(Registrations_i) = \alpha + \beta \ln(Pop_i) + \varepsilon_i \quad (4)$$

The regression coefficient β estimates the elasticity of Pop and *Registrations*, α is a constant and ε_i an identical and independently distributed error term.

Since the price for a domain registration is constant, the elasticity of registrations with respect to Pop should equal 1 by theory if the supply of domains was effectively unconstrained. The regres-

sion coefficient β would then have a value of 1. A coefficient estimate significantly below 1, however, would support the hypothesis of virtual space scarcity:

H1 – The relative increase in the number of registered domains is smaller than the relative increase in the number of potential domain registrants.

Defining E at the domain level also allows accounting for any differences in linguistic quality between domains and the resulting differences in commuting costs and registrations. Zipf (1936) shows that shorter words tend to appear more frequently in natural languages than long expressions. Similarly, the length of Twitter hashtags is inversely related to their usage frequencies (Cunha et al. 2011). If his principle of least effort also holds true in domain space, shorter domains will be registered more often than long domain names. For instance, bearers of long surnames are less likely to register a domain containing their name than somebody with a relatively short name. To give a simplifying example, domains derived from the keywords “Pennington Associates Milwaukee” might be more tedious to type than any from “Carr Associates Miami”, making the former less likely to appear in registrations.

If domain length is a valid proxy for the effort required by users to access a virtual location, then an inverse relation between the length of a string and the number of registrations containing this string can be expected.

H2 – The likelihood of a character string being registered as a domain name decreases in the length of the string.

To test H2, the variable domain *Length* is added to (4):

$$\ln(\text{Registrations}_i) = \alpha + \beta_1 \ln(\text{Pop}_i) + \beta_2 \ln(\text{Length}_i) + \varepsilon_i \quad (5)$$

A negative estimate for the regression coefficient β_2 can be interpreted as evidence for different levels of effort required by users – or for the equivalent to commuting costs required by the monocentric city model.

In a similar fashion, the number of keywords within a domain name can be interpreted as an additional measure for commuting costs, as more keywords require more effort when memorizing and recalling. However, combining multiple keywords results in a trade-off between brevity and descriptiveness. In case the domains “pizza.com” or “pizzaboston.com” are already taken, “tastypizzaboston.com” might still be available as the electronic storefront of a local pizza place. Theoretically, each additional keyword increases the number of potential domain names by several orders of magnitude: If the total number of viable single keywords is W , then W^2 two-keyword combinations, or W^3 three-keyword combinations are possible. Whoever is willing to accept the higher effort required to access a domain consisting of many keywords has plenty of choice. This trade-off between availability and domain quality reconciles the view of seemingly unconstrained domain supply and the observation that short, low-effort domains are not easy to come by: Just add a few more keywords and you can have any domain you want.

This notion motivates one last hypothesis:

H3 – Domain space is less constrained for combinations of multiple keywords than for single-keyword domains.

To test H3 empirically, the domain registrations for each surname or MSA i are further subdivided into 4 subgroups k , where k denominates the number of keywords in each name. For instance, $Registrations_{Boston,2}$ counts the number of COM registrations containing BOSTON and one additional keyword, $Registrations_{Boston,3}$ is the number domains with two additional keywords, and so

forth. The dummy variables D_n are defined as 1 if $k = n$ and 0 otherwise. All β 's are regression coefficients:

$$\ln(\text{Registrations}_{i,k}) = \alpha + \beta_1 \ln(\text{Pop}_i) + \sum_{k=3}^{k=5} \beta_{2,k} D_{i,k} \ln(\text{Pop}_i) + \sum_{k=2}^{k=5} \beta_{3,k} D_{i,k} + \beta_4 \ln(\text{Length}_i) + \epsilon_i \quad (6)$$

DATA

The domain name system was designed in a distributed fashion⁶ with as little information as possible managed in centralized registries. Each Top Level Domain, for instance, administers separate databases for its domain registrations and delegates the actual task of managing the information linked to a domain to a large number of decentralized domain name servers. So called *TLD zone files* keep track of the authoritative domain name servers for each domain under a specific TLD. The zone file for the COM domains can be downloaded from Verisign (2014b). Strictly speaking, this zone file does not contain *all* registered domains, but only active COM domain names with a DNS entry, which account for more than 99.5% of all domains (Verisign 2014a).

Domain names are often comprised of multiple keywords linked together into one character string, complicating any analysis of the names' meaning. This concatenation is reverted and all domains are split into their base keywords, employing an automatic programming interface described in Huang et al. (2010) and hosted by Microsoft Research (Microsoft Research 2014).⁷ The next step identifies groups of domain names that contain popular surnames or city names as a keyword. This segmentation builds on the premise that the number of domains per city resident or bearer of a surname is (on average) the same across all cities or names: Why would there be a different number of domain registrations domains containing "Miller" versus "Smith", for example, after accounting for the total number of citizens named Miller or Smith and the length of the name?

⁶<http://tools.ietf.org/html/rfc1034>

⁷The list of domains and the corresponding keywords are available from author on request.

Similarly, the demand for domains is also expected to be equal across cities after controlling for the size and the socio-economic composition of the city populations. The cities' average income per capita and the share of the population holding university degrees are added to the regression equation as control variables since Goldfarb and Prince (2008) found high-income and well educated people to be overrepresented among early Internet users. Beyond education and income, why would there be higher or lower per capita demand for domains in e. g. Houston versus Dallas?

The US Census (US Census 2014b) provides an overview of the most popular surnames from the year 2000, including frequency counts and basic demographic information. In addition, this source also lists the population numbers of all US Metropolitan Statistical Areas (MSA) (US Census 2014a). Before linking population numbers and domain registrations, MSAs with the same name (like Portland Oregon and Portland Maine) are aggregated into one observation and the corresponding population numbers are added up.

===== Insert Table 1 about here =====

Table 1 displays summary statistics for all domain names and for subsamples of domains containing MSA names or popular surnames. Out of 107.5 million domain names, 9.6 million domains contain a popular US surname and 2.98 million domains feature a city or MSA name⁸. Overall, surname domains are on average 14.65 characters long (not counting the “.com”), which is 1.3 characters more than the average length of a COM domain. Surname domains also contain more keywords with an average of 2.67 words versus 2.47 for all domains. City related domains are even longer with 18.78 characters and 3.09 keywords on average.

⁸Moss and Townsend (1997) used a different specification when calculating the geographic domain registration density by dividing the number of domains registered by residents of a city over the city's total population.

As demand for a certain group of domains increases, registration of domains in these groups also increases, including an increasing number of long names. For the 500 most frequent surnames in the US census, 60% more domains have been registered than for the following 500 surnames in the frequency ranking (5.93 million vs. 3.7 million). The higher registration numbers come at a cost. The more sought after domains are on average longer and contain more keywords than domains containing less frequent surnames. All differences in length and keyword counts are statistically significant with t-values above 2.6.

For cities, the differences become even more pronounced. When splitting the sample of MSAs at the population median, the most populous cities account for more than 4 times the number of domain registrations originating from the lower half of MSAs (2.49 million vs. 0.6 million) and the magnitude of the average length-differences is also substantially bigger.

The variable *Length* captures the number of characters for each surname or MSA name in the sample. On average, a top-1000 surname is 6.08 characters long. The maximum is 11 and the minimum is only 2 characters (US Census 2014b). For MSA names, the mean of *Length* is 8.67, and the range is 4 to 16 characters (US Census 2014a).

For the MSAs, the share of population having attained a bachelor or graduate degree ranges from 7% (Merced, CA) to 34% (Boulder, CO), and the minimum income per capita is USD 14,126 (McAllen, TX) while the maximum value is USD 48,900 (Bridgeport, CT).

RESULTS

Figure 3 presents the population numbers per MSA (a) and the frequencies of surnames (b) plotted against the number of corresponding domain registrations. The logarithms of the domain registration numbers are in both cases a linear function of the logarithms of the underlying demand met-

rics. However, the estimated trend lines (solid lines), have a slope below one (dashed lines) indicating constrained markets. Panel (a) in Table 2 quantifies the magnitude of the elasticities: If the population of a MSA is one percent bigger than the population of another MSA, the difference in domain registrations is only about 0.80% higher. For surnames, the elasticity is even lower. A one percent increase in surnames pushes domain registrations up by only 0.74%. The estimated coefficients for $\ln(Pop)$ are statistically different from both 0 and 1 at the 1% confidence level, confirming Hypothesis 1.⁹ Domain demand is linked to the knowledge economy: Higher levels of educational attainments in an MSA lead to higher levels of domain demand. After controlling for education, the coefficient for income becomes insignificant.

The estimate of 0.80% for MSAs may be a conservative figure as the market may be in fact even more constrained due to two reasons. First, some MSA names undoubtedly have marketing appeal to users from other parts of the world. Global brand names like “New York” or “Los Angeles” are not exclusive to residents from these MSAs only. Any non-native usage could inflate domain registration numbers for larger MSAs and bias the elasticity estimates upwards. Second, large MSAs like New York could have disproportionately more small businesses and retail shops given their consumption variety. On other hand, smaller MSAs might depend more heavily on branches of larger chains that do not require their own web presences. Analyzing the link between the industrial composition of MSAs and domain registrations could be an interesting aspect for a future study.

==== Insert Figure 3 about here =====

===== Insert Table 2 about here =====

⁹The results are robust with regard to the population size of MSAs or the frequencies of the surnames: When splitting the sample at the median of Pop all estimated elasticities are estimated to remain significantly smaller than 1.

Longer domains are less desirable than shorter names. Panel (b) of Table 2 lists the estimated regression coefficients from (5). The coefficients for $\ln(\text{Length})$, β_2 , are negative and significantly different from 0 for surnames and MSA names alike. Increasing the length of a surname from the median (6 characters) to the 75th percentile (7 characters), reduces the number of domain registrations by a staggering 24%¹⁰. For MSA names, adding one more letter to the median of 9 results in an 14% lower number of registrations. H2 is clearly confirmed.

H3 hypothesizes that registrants circumvent the problem of their desired domain already being registered by adding more keywords. The regression coefficients from (5) are displayed in Panel (c) of Table 3. The negative coefficients for $D_3 \text{ Keywords}$, $D_4 \text{ Keywords}$, and $D_5 \text{ Keywords}$ confirm that domains with fewer keywords are more popular than longer alternatives. The base case, $D_2 \text{ Keywords}$, comprises of the shortest and most sought-after group of domains consisting of only the MSA name or surname in conjunction with one more keyword. The clear preference for less complex names decreases for larger values of Pop , due to the positive estimates for the interaction terms' $D_k \ln(Pop)$ coefficients. For example, in all MSAs the population is larger than 55,000, so the -0.436 for $D_3 \text{ Keywords}$ is fully offset by 0.132 times $\ln(55,000)$.

The estimated elasticities for two-keyword MSA domains is 0.667, which is below the overall elasticity of 0.795 estimated in model (a) for all keyword lengths before. The elasticity for three-keyword MSA domains is already higher at 0.799 (0.667+0.132), while four-keyword domains and five-keyword domains are basically unconstrained. For surnames, the elasticities also increase each time a new keyword is added. These estimates support Hypothesis 3: adding more keywords to a domain name reduces supply constraints.

¹⁰ $1 - \exp((\ln(7) - \ln(6)) * -1.787) = 0.24$

CONCLUSION AND DISCUSSION

This study investigates whether the primary market for Internet domain names can be analyzed using standard urban economic theories. It shows that, indeed, core prerequisites are met which allow the application of the monocentric city model to virtual space: an equivalent to commuting costs can be constructed from differences in linguistic properties of domain names and the supply of “central” virtual locations that exhibit low commuting costs is constrained.

The empirical results show that the number of registered domain names has been increasing at a slower pace than than fundamental demand for domain space. The magnitude of this effect is large in economic terms: If the elasticities found for domains containing surnames or city names are a valid estimate for the overall elasticity, then the total number of domain names could be up to a quarter higher if more domains were available for registration.

The introduction of new global top level domains has the potential to serve some of the demand not met by the current domain extensions (additional benefits of more competition in TLD space, like technical innovation, lower prices, joint marketing efforts and overall more choice for consumers have been described by Mueller [1998] and others). It is too early to tell whether the new space will be accepted as a viable alternative to the established space. Replacing the ubiquitous COM or the country-specific TLDs like NL, DE, or CO.UK with a new TLD is comparable to adding one more keyword to a name: It increases choice massively but additional keywords are only viable in very crowded segments of domain space (see Table 3, Panel [c]).

In land markets, external factors like the topography of a metropolitan area can exacerbate land scarcity (Saiz 2010): cities that are physically constrained by water or mountainous terrain exhibit steeper land price gradients than places in open landscapes. Analogous policy choices in the ad-

ministration of the web's address system can partially remove existing constraints on domain supply. Launching additional top level domains will alleviate scarcity in web locations, but not fully overcome it. Figuratively speaking, the new TLDs flatten hills and fill in water, but once these obstacles are removed, the overarching constraints will kick in again: The set of catchy keywords that appeal to humans is still bound by the way we process language, even if we had unlimited choice in top level domains. Legend has it that Mark Twain advised to buy land, since "they have stopped making it"¹¹. Similarly, one can argue that investing into (top level) domains is a promising business venture, since we have stopped inventing new languages, at least at a large scale.

Additionally, the paper confirms old market wisdom: Longer names are indeed registered less frequently than shorter names. In follow up studies, the virtual equivalent to commuting costs can be extended to analyze the effect of other linguistic characteristics like the keyword types, special characters, numerals or hyphens. Additional keywords in a name might provide more choice of possible domains but they come at the cost of longer overall names.

These findings can be generalized to other location systems based on natural language like Twitter handles, identifiers in online communities or names of existing companies, individuals or even cities. Latino and Hispanic immigrants display a strong preference for places with Spanish names in the south-western USA, after accounting for county fixed effects and other locational variables (Saiz 2014). Since city names influence residential choices, it would not be surprising if a city bearing a short, tongue-friendly name attracts more new residents than less appealingly named alternatives: Why settle with a linguistic challenge like Schenectady if Albany is so close by?

Finally, by applying the monocentric city model to virtual space, one can transfer long-established findings from traditional land markets to domain markets: For instance, the intensity of space use

¹¹ Multiple versions of this quote are widely distributed – a source has not been handed down, however.

is predicted to be higher for domain names with low commuting costs than for locations in the linguistic periphery – a prediction which could be tested in follow up studies.

More research is also needed to understand the determinants of domain registrations in time. The current study draws its conclusion from one cross-section only. Technological change could reduce the effort required to navigate to a website and might channel more demand to peripheral locations. Examples of relevant technological change include the auto-complete function in the browser's address bar which automatically fills in any long domain names, in case they have been visited in the past. Additionally, it would be interesting to investigate where unsuccessful registrants turn after not having found a suitable name. The substitution between “owner-occupied” domain space and “rented” locations on e. g. social media platforms is not understood in academic literature yet.

REFERENCES

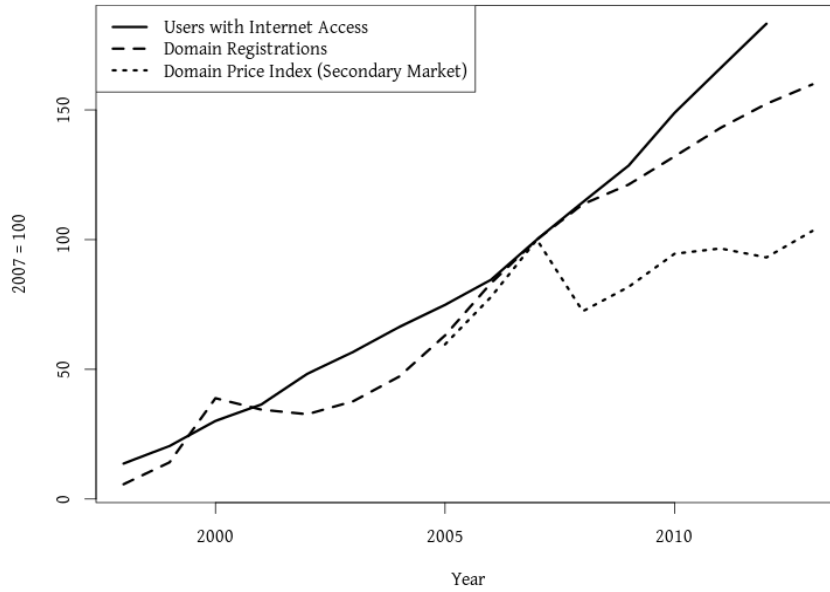
- Alonso, W. (1964). *Location and land use: toward a general theory of land rent*. Cambridge, MA: Harvard University Press.
- Batty, M. (1997). Virtual geography. *Futures*, 29(4/5), 337–352.
- Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral cortex*, 19(12), 2767–96. doi:10.1093/cercor/bhp055
- Cairncross, F. C. (1997). *The death of distance: How the communications revolution is changing our Lives*. Cambridge MA: Harvard Business Review Press.
- Cohen, G., & Burke, D. M. (1993). Memory for proper names: a review. *Memory*, 1(4), 249–63. doi:10.1080/09658219308258237
- Cunha, E., Magno, G., & Comarela, G. (2011). Analyzing the dynamic evolution of hashtags on twitter: a language-based approach. In *Proceedings of the Workshop on Language in Social Media* (pp. 58–65). <http://dl.acm.org/citation.cfm?id=2021117>. Accessed 6 October 2014
- Damasio, H., Tranel, D., Grabowski, T., Adolphs, R., & Damasio, A. (2004). Neural systems behind word and concept retrieval. *Cognition*, 92(1-2), 179–229. doi:10.1016/j.cognition.2002.07.001

- Friedman, T. (2005). *The world is flat: A brief history of the twenty-first century*. New York: Farrar, Straus and Giroux.
- Gaspar, J., & Glaeser, E. L. (1998). Information technology and the future of cities. *Journal of Urban Economics*, 43, 136–156. doi:10.1006/juec.1996.2031
- Geltner, D. M., Miller, N. G., Clayton, J., & Eichholtz, P. (2001). *Commercial real estate: Analysis & investments*. Mason, OH: Thomson.
- Goldfarb, A., & Prince, J. (2008). Internet adoption and usage patterns are different: Implications for the digital divide. *Information Economics and Policy*, 20(1), 2–15. doi:10.1016/j.infoecopol.2007.05.001
- Huang, J., Wang, K., Behr, F., Gao, J., Miao, J., & Li, X. (2010). Exploring Web Scale Language Models for Search Query Processing Categories and Subject Descriptors. In *Proceedings of the 19th international conference on WWW 2010* (pp. 451–460). Raleigh, NC.
- Humphreys, G. W., & Forde, E. M. E. (2001). Hierarchies, similarity, and interactivity in object recognition: “Category-specific” neuropsychological deficits. *Behavioral and Brain Sciences*, 24, 453–509.
- Ihlanfeldt, K. R. (1995). The importance of the central city to the regional and national economy : A Review of the arguments and empirical evidence. *Cityscape*, 1(2), 125–150.
- Internet Corporation for Assigned Names and Numbers, ICANN. (2013). Annual report 2013. <http://www.icann.org/en/about/annual-report/annual-report-2013-en.pdf>
- Internet Corporation for Assigned Names and Numbers, ICANN. (2014). Monthly Registry reports. <https://www.icann.org/resources/pages/reports-2014-03-04-en>
- Internet Corporation for Assigned Names and Numbers. (2014). New gTLDs: Program statistics. <http://newgtlds.icann.org/en/program-status/statistics>
- Lindenthal, T. (2014). Valuable words: The price dynamics of internet domain names. *Journal of the Association for Information Science and Technology*, 65(5), 869–881.
- Lindenthal, T., & Loebbecke, C. (2014). Pricing quality attributes of Internet domain names: A hedonic model for words. In *Twentieth Americas Conference on Information Systems*. Savannah.
- Microsoft Research. (2014). Microsoft Web N-gram services. <http://web-ngram.research.microsoft.com/>
- Mills, E. (1972). *Studies in the structure of the urban economy*. Baltimore, MD: The John Hopkins University Press.
- Moss, M. L., & Townsend, A. (1997). Tracking the net: Using domain names to measure the growth of the internet in U.S. cities. *Journal of Urban Technology*, 4(3), 47–60.
- Mueller, M. L. (1998). The battle over Internet domain names. Global or national TLDs ? *Telecommunication Policy*, 22(2), 89–107.
- Muth, R. (1969). *Cities and housing: the spatial pattern of urban residential land use*. Chicago: University of Chicago Press.

- Pascal, A. (1987). The Vanishing City. *Urban Studies*, 24(6), 597–603.
- Saiz, A. (2010). The geographic determinants of housing supply. *The Quarterly Journal of Economics*.
- Saiz, A. (2014). Hispanics and recent latino immigrants locate disproportionately in places with Spanish names. MIT Working Paper Series. [http://web.mit.edu/urbeconlab/5 Immigration and Hispanic Growth in the US/Project 1_Immigrants and Spanish City Names/Immigrants and Spanish City Names.pdf](http://web.mit.edu/urbeconlab/5%20Immigration%20and%20Hispanic%20Growth%20in%20the%20US/Project%201_Immigrants%20and%20Spanish%20City%20Names/Immigrants%20and%20Spanish%20City%20Names.pdf)
- The Economist Online. (2010, October). Sex.sells. http://www.economist.com/blogs/dailychart/2010/10/domain-name_prices
- Tranos, E., & Nijkamp, P. (2013). The death of distance revisited: Cyber-place, physical and relational proximities. *Journal of Regional Science*, 53(5), 855–873.
- US Census (2014a). Annual estimates of the population of metropolitan and micropolitan statistical areas: April 1, 2010 to July 1, 2012. <https://www.census.gov/popest/data/metro/totals/2012/tables/CBSA-EST2012-01.xls>
- US Census (2014b). File A: Top 1000 Names. <http://www.census.gov/genealogy/www/data/2000surnames/Top1000.xls>
- Verisign. (2014a). Domain Name Industry Brief. <http://www.verisigninc.com/assets/domain-name-report-april2014.pdf>
- Verisign. (2014b). Zone File Information. http://www.verisigninc.com/en_US/channel-resources/domain-registry-products/zone-file-information/index.xhtml
- Zipf, G. K. (1936). *The psycho-biology of language*. London: Routledge.
- Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Cambridge: Addison-Wesley.

TABLES AND FIGURES

Figure 1: Internet Usage, Domain Name Registrations, and Re-sale Prices



Notes: Domain registrations and the worldwide number of Internet users grew at similar rates up to 2007. Later, domains get registered at a slower pace. The prices paid for already existing domains increased by 60% from 2006 through 2013. Data: (ICANN (International Corporation for Assigned Names and Numbers) 2014), (Lindenthal 2014), Zooknic.com, Worldbank.

Figure 2: Demand and Supply for Domain Names

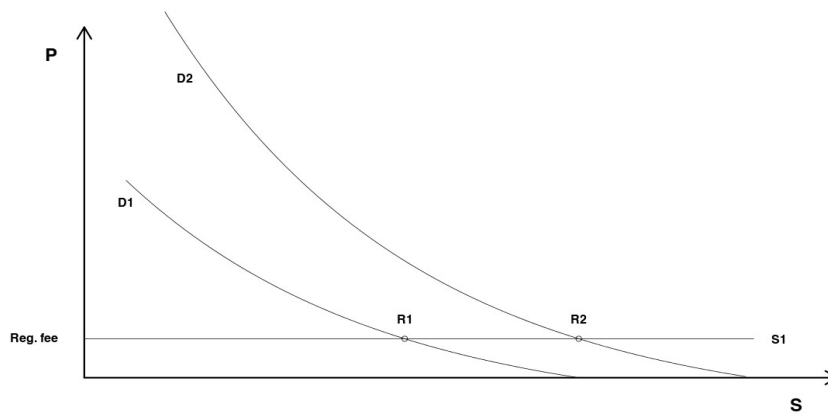
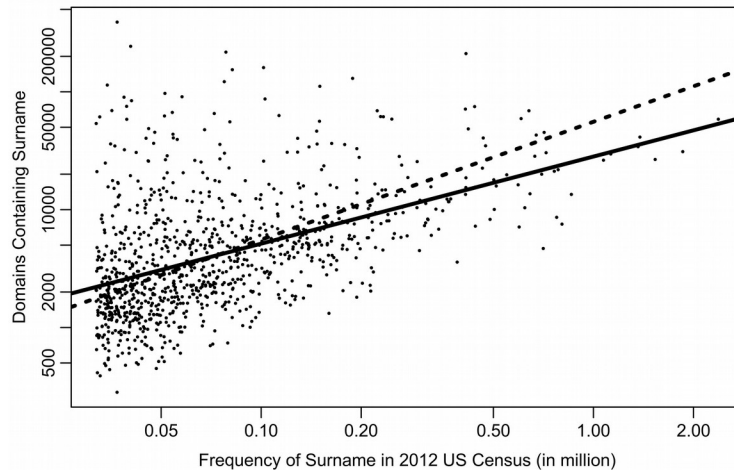
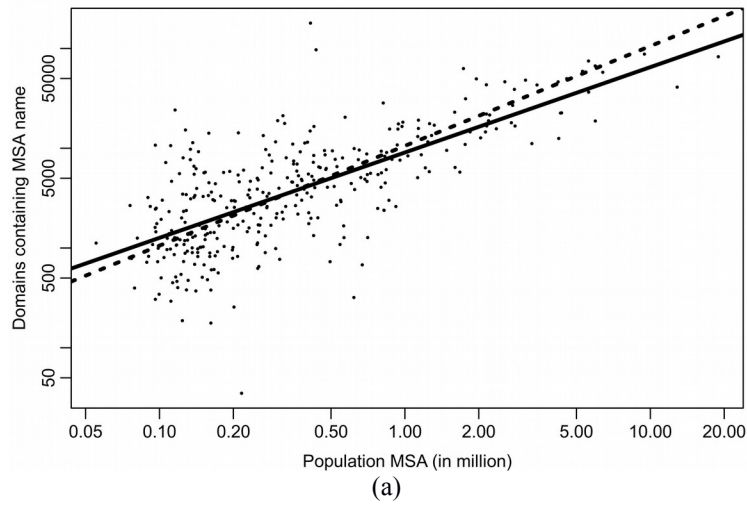


Table 1: Summary Statistics

	<i>All COM domains</i>	<i>COM domains containing one of 1000 most frequent US surnames</i>				<i>COM domains containing US MSA names</i>			
		all	Top 50%	Bottom 50%	t(Δ)	All	Top 50%	Bottom 50%	t(Δ)
Registrations (in million)	107.46	9.62	5.93	3.69		2.96	2.49	0.47	
Characters/SLD, average	13.30	14.65	14.85	14.45	3.77	18.78	19.19	18.37	3.10
Keywords/SLD, average	2.47	2.67	2.72	2.63	6.46	3.09	3.11	3.07	0.69

Notes: Domains containing popular surnames or city names account for a large share of all registered COM domains. Overall, the number of characters and keywords per domain is larger for more frequent (Top 50%) surnames and more populous cities. All but one difference (Δ) in length and number of keywords are statistically significant, with t-values above 2.6.

Figure 3: Domain registrations and fundamental demand



Notes: All axes in logarithmic scales. The solid lines represent the estimated elasticity between the population of US MSAs (a) or the frequency of surnames in the 2012 US Census (b) and the number of corresponding registrations of COM domains. For both panels, the actual elasticities are lower than the unconstrained elasticities shown as dashed lines. Domain registrations grow slower than fundamental demand for cyberspace.

Table 2: Regression Coefficient Estimates

<i>Variable</i>	(a)		(b)		(c)	
	<i>MSA</i>	<i>Surnames</i>	<i>MSA</i>	<i>Surnames</i>	<i>MSA</i>	<i>Surnames</i>
<i>Constant</i>	6.253*** (6.135)	0.036 (0.481)	8.414 (5.405)	3.188*** (0.466)	9.787* (5.049)	4.253*** (0.519)
<i>ln(Pop)</i>	0.795 ** *	0.739 ***	0.775 ***	0.741 ***	0.667 ***	0.622 ***
<i>ln(Length)</i>	(0.051)	(0.043)	(0.045) -1.476 *** (0.149)	(0.038) -1.787*** (0.107)	(0.076) -2.001*** (0.138)	(0.045) -2.063*** (0.064)
<i>ln(Mean income)</i>	-0.643 (0.558)		-0.503 (0.491)		-0.609 (0.453)	
<i>ln(% with degree)</i>	0.890 ** *		0.995 *** (0.286)		0.994 *** (0.264)	
<i>D₃ Keywords (vs. 2)</i>					-0.436 (1.338)	-3.031 *** (0.716)
<i>D₄ Keywords (vs. 2)</i>					-1.950 (1.338)	-4.622 *** (0.716)
<i>D₅ Keywords (vs. 2)</i>					-3.274 (1.338)	-5.919 *** (0.716)
<i>D₃ Keywords ln(Pop)</i>					0.132 (0.105)	0.246 *** (0.064)
<i>D₄ Keywords ln(Pop)</i>					0.190 * (0.105)	0.266 *** (0.064)
<i>D₅ Keywords ln(Pop)</i>					0.198 * (0.105)	0.271 *** (0.064)
<i>Adj. R2</i>	0.52	0.23	0.62	0.40	0.45	0.64

Notes: Coefficients are estimated with ordinary least squares regressions based on Equations (4), (5), and (6). The dependent variable is the natural logarithm of domain registrations. Standard errors in parentheses. Coefficients that are statistically different from 0 at 1%, 5%, or 10% confidence levels are marked with ***, **, or *, respectively.