Technology Diffusion in the Mobile Market – A Case Study

Tim Schulz

Summer 2016

Abstract

I analyze the development of mobile internet usage from the beginning of 2008 to the end of 2015 based on the S-curve model of technological diffusion proposed by Griliches (1957) using Wikimedia page traffic as a case study. I extend the model by allowing for temporary changes due to the introduction of various smartphones to the market. These changes are estimated to be positive as one might expect and the effect of new devices on mobile internet usage is diminishing over time.

This analysis is done for every language that Wikimedia pages are written in. While the numerical estimation does not converge to a unique solution for some of the more exotic languages, it is quite reliable for more popular languages, the results for some of which are shown in this paper as examples.

1 Introduction

Smartphones have gone from virtual non-existence to near ubiquity within less than a decade. Between 2007, the year the first iPhone was released, and 2015, the number of smartphones sold globally increased by a factor of more than 10.[1] A person can only use so many smartphones at once though and, for the first time in years, sales growth is expected to fall below double digits.[2] However, it would be premature to conclude that the importance of smartphones and other mobile devices, such as tablets, has reached its peak. Therefore, merely measuring growth through sales will soon be difficult and only provide limited insight.

Alternatively, growth could be measured by usage. It is impractical (if not impossible) to record everybody's smartphone usage. Instead, by using Wikipedia as a globally representative website (at least for countries where it is not banned), this can be approximated.

---

[1] http://www.statista.com/statistics/263437/global-smartphone-sales-to-end-users-since-2007/
[2] https://www.idc.com/getdoc.jsp?containerId=prUS41061616

Since the end of 2007, the Wikimedia Foundation has been publishing hourly page traffic for all its websites including information on whether an article was accessed in its traditional (desktop/laptop) format or as a mobile site and what language the article is published in.

I use this data to describe the share of the number of mobile visits relative to traditional web traffic. The underlying model is that of a technology S-curve as first proposed by Griliches (1957). Additionally, I allow for temporary increases of mobile usage due to the introduction of a new device. The magnitude of this boost in the share of mobile traffic across several generations of mobile devices quantifies the process of technological diffusion.

I find that the S-curve model describes the development of relative mobile internet usage on Wikimedia pages well. There is evidence in favour of a positive effect of new smartphones introduced to the market even though that effect is diminishing with each generation.

2   Literature Review

Processes revolving around technology have been the subject of interest in economic research for some time. However, in most cases "technology" is considered in the context of production usually as a factor in the output function that increases productivity. Technology in the more colloquial, every-day sense of the word is not the same. From here on, "technology" will refer to this use of the word in the sense that it describes hardware (usually in the form of electronics) and software produced by the technology (or high-tech) industry.

Innovation, the introduction of new technology, is not the immediate subject of this paper. Rather, I will focus on the process that leads from innovation to widespread adoption of technology: technology diffusion. This process was first quantified by Griliches (1957)[1] using the cumulative logistic distribution function in the description of the adoption of hybrid corn in the US. This method not only matches the often observed S-curve of technological adoption over time,[3] but also offers the mathematical convenience of being able to identify a unique "rate of acceptance".

This concept of technology diffusion has since been applied to many areas, including computers (Caselli and Coleman (2001)[2]) and hard drives (Christensen (1992)[3]). While the latter of these two papers is specifically written in the context of firms where new technology results in a higher number of bits written on a given area of a hard drive, it can be easily translated into the context of consumer electronics, where instead of a measure of

---

[3]Initial incorporation of the new technology is slow, then speeds up and eventually slows down again as the "market" (whatever the context) gets close to being saturated.

hard drive performance, the dependent variable in this paper will be mobile internet traffic. Just as the technological concept of a hard drive had already existed for a long time but the "output" (bit density) kept improving, so has the concept of mobile internet access existed for some years but the adoption of this technology is still in progress.

More related work has been done by Czernich et al. (2011)[4] who analyse broadband penetration rates over time based on the S-curve model. Majumdar and Venkataraman (1998)[5] analyse technological adoption and diffusion in the context of the telecommunications industry. In this case, technology diffusion is not just a function of time and knowledge (and other variables) but it is also specifically subject to network effects. Hall and Khan (2002)[6] extend this idea by introducing the concept of indirect network effects using the example of DVD players. Even if the technology exists and people are, in principle, prepared to adopt it, only once enough people are interested in buying DVDs is it worthwhile for production companies to switch over from video cassettes. In turn, only then might some people be willing to invest in a DVD player. The same can be argued to be true for mobile internet. While Wikipedia articles did not become available in mobile formats one by one[4], there are still network effects associated with other websites. If a user knows from past experiences that most websites are poorly formatted for mobile devices, he is less likely to access the internet from a smartphone even if the the website in question has a mobile version. If, however, (almost) all major websites have mobile versions, a user is much more likely to access the internet using a mobile device than was the case before. Mobile internet usage can therefore be argued to be a function of something best described as the "overall mobile internet experience". Due to these network effects, Wikipedia page traffic can be seen not only as representative for the rest of the internet but also as a function of mobile services provided by other website operators.

3   Data Description

The data underlying this paper are hourly Wikimedia page view statistics.[5] This database contains the number of visits and the amount of data (in bytes) transmitted in response for every page of the various Wikimedia projects accessed within a given hour from 18:00 (UTC), December 9, 2007 onward. These statistics are categorised by language of the page,

---

[4]The website code only has to be adjusted once to redirect visitors with mobile devices to a differently formatted version of the article (regardless of the article itself).

[5]The root directory of the data dumps can be found here:
https://dumps.wikimedia.org/other/pagecounts-raw/

page or project type[6], page/article name, number of visits [7], and the response size.

Since the individual pages are not of interest in this analysis, the statistics are first aggregated to hourly language-type data by summing the number of visits and the amount of traffic within each language-type category. Additionally, the number of pages that are aggregated to one data point is recorded.

A note regarding the "mobile" type is in order: This category does not differentiate between pages. It is already an aggregation of all mobile Wikimedia pages within one language. As a consequence, the "mobile" category of each language seemingly contains only one page after aggregation, meaning the total number of pages is only available for non-mobile categories. However, the number of visits and amount of traffic are still analogous to other categories after aggregation. Therefore, the different non-mobile types are added up to one category in order to retain comparability with "mobile". The resulting panel contains the hourly number of visits and the amount of data of all mobile and non-mobile parts of the Wikimedia Project by language. There are (roughly) 70,000 hours and about 200 languages for a total of 14 million observations per variable. However, since hourly data is extremely sensitive and technological adoption is probably measured in weeks if not months or years rather than hours, the data is further aggregated to weekly frequency. This has the added benefit of not having to worry about time fixed effects (possibly by language) of frequencies higher than that.

Lastly, a time series for both the number of visits and the amount of traffic is calculated. This time series contains the ratio, R_{lt} of mobile activity relative to the sum of mobile and non-mobile activity per hour h in week t and language l. Specifically, this is the ratio of the sums such that

$$R_{lt} = \frac{\sum_{h \in t} m_{lh}}{\sum_{h \in t} m_{lh} + \sum_{h \in t} n_{lh}},$$

where m and n denote mobile and non-mobile traffic summed up within a specific week. Summing before dividing gives less weight to possible outliers and incorrectly recorded data than a weekly mean of hourly ratios would.

The main explanatory variable, besides time, is the date of introduction of new mobile technology. In the case of iPhones, this is relatively straight forward given the small number of devices and the low frequency of new releases. In this case, the release date of a new

---

[6]The types are Wikibooks, Wiktionary, Wikimedia, Wikipedia, mobile, Wikinews, Wikiquote, Wikisource, Wikiversity and Mediawiki.

[7]The number of visits is the total number, not unique visits. I.e. if a person accesses a Wikipedia article multiple times within hone hour, each of these events is counted. It is not how many *people* visited a page.

generation of the iPhone is used. For Android phones, however, this is not feasible given the large number of devices and irregular release dates across manufacturers. Therefore, the release dates of the Samsung Galaxy S lineup is used instead as a representative for Android devices.[8]

The data together with the Python scripts that were used to download and aggregate the hourly data files can be found at `https://github.com/timschulz91/ma-project`. Also included here are the R scripts that were used for modelling and analyzing the data.

4   Model and Methods

Since the first smartphones only appeared towards the end of of the 2000s, the ratio of mobile to total activity will be equal to or close to zero in the beginning of the time series and can naturally not exceed 1. That is, $R_{lt} \in [0, 1]$. Given these restrictions, the development of R_{lt} is assumed to follow a logistic growth function which resembles the observed S-curve and has the convenient property of a simple derivative which can be interpreted as the instantaneous growth rate at a given time. The S shape is quite plausible in the context of technological adoption. Any time before the new technology is introduced, the rate of adoption is necessarily zero. In early stages, few people will use the technology and growth will be minimal. These two properties are well represented by the logistic curve as it converges to zero "to the left". After some time, growth accelerates but eventually slows down as the market becomes saturated. This is captured by the convergence to 1 in the logistic function as the underlying variable (in this case time) increases.

Realistically, mobile traffic will likely never represent all traffic. In other words, the ratio of interest will never equal 1 and probably also never approach it. Instead, it makes more sense to assume convergence to some other upper bound K_l that can be different for each language.

$$\lim_{t \to \infty} R_{lt} = K_l < 1$$

This closely follows the initial work on technology S-curves done by Griliches (1957) where

$$R_{lt} = K_l \Lambda \left(\alpha_l + \beta_l t \right) + \epsilon_{lt}$$

---

[8]The various generations of the Samsung Galaxy S have the most consistent large number of sales of any family of Android devices.

and

$$\Lambda(x) = \frac{1}{1 + e^{-x}},$$

which allows for different "starting points" α (in terms of Griliches, the first time R exceeds 10%) and rates of acceptance/adoption β for each language.

Therefore, when fitting the logistic function to the use of mobile websites within one language, K_l, α_l and β_l have to be estimated simultaneously. This is achieved using the `nlstools` package for non-linear models in R. `nlstools` provides functions for the numerical minimization of the sum of squared residuals resulting from the estimated object (the "nls" part stands for nonlinear least squares). For the purpose of this paper, the PORT method is used as it allows to impose constraints on some of the parameters ($0 \leq K < 1$, $0 \leq \beta$).

If, indeed, the introduction of new mobile technology results in a temporary acceleration of mobile internet usage, one would expect a series of positive residuals after the introduction date. The size of these residuals can be estimated using dummy variables that indicate whether a week t falls between the release dates of a device and its successor. More formally, assuming introduction occurs at time t_j and the successor device is introduced at time t_{j+1},

$$
\begin{aligned}
R_{lt} = K_t \Lambda\left(\alpha_l + \beta_l t\right) &+ \delta_1 \mathbb{1}\left\{t \in [t_1, t_2)\right\} \\
&+ \delta_2 \mathbb{1}\left\{t \in [t_2, t_3)\right\} \\
&+ \delta_3 \mathbb{1}\left\{t \in [t_3, t_4)\right\} \\
&\vdots \\
&+ \epsilon_{lt} \\
= K_t \Lambda\left(\alpha_l + \beta_l t\right) &+ \sum_{j=0}^{J} \delta_j \mathbb{1}\left\{t \in [t_j, t_{j+1})\right\} + \epsilon_{lt}
\end{aligned}
$$

so that δ_j measures the effect of of generation j of the technology on the ratio of mobile to total traffic.

In the context of more than one device lineup, this can be extended to

$$R_{lt} = K_t \Lambda\left(\alpha_l + \beta_l t\right) + \sum_{j=0}^{J} \delta_j \mathbb{1}\left\{t \in [t_j, t_{j+1})\right\} + \sum_{g=0}^{G} \gamma_g \mathbb{1}\left\{t \in [t_g, t_{g+1})\right\} + \cdots + \epsilon_{lt},$$

where δ_j and γ_g measure the effect of the jth and g^{th} generation of two separate lineups of

6

devices, respectively.

Theoretically speaking, one could have t_j depend on the language. This works in cases where a language is almost exclusively spoken within one country (e.g. Danish) but not in cases where a language is spoken in several countries with possibly different release dates of new devices (e.g. English). However, usually devices are introduced worldwide if not on the same day, then in the same week. Therefore, having t_j vary by language is not necessary. Furthermore, for the remainder of the paper, the model is estimated separately for each language so that any language subscript l is dropped.

5 Results

After removing observations for niche languages such as Pirate and others for which there were no data for more than 40% of the recorded time range, 200 unique languages remain. Some of these languages are used very little and as a consequence, even after aggregating to weekly ratios, the data is very noisy. For many of these, the minimization in the estimation procedure of the model presented above did not converge. This was the case for 49 languages.

Also, in some cases the maximization "ran into" the upper bound on K, which is 1. Since a weekly rate of adoption of 1 is highly unlikely, these cases were removed as well under the assumption that it is not possible to solve numerically for a unique minimum. 134 languages remained.

Throughout the paper I use some of the ten most prominent languages on Wikipedia as example results. Since there are at least 1 million Wikipedia articles alone for each of these languages,[9] they should be the least prone to random fluctuations and missing data and therefore deliver the most reliable estimates of the S-curve.

5.1 General Fit of the S-Curve

Figure 1 shows the observed ratio and fitted S-curve for the English language (the language with the most articles) and the expected upward trend is obvious. The rate of acceptance/adoption visibly starts decreasing in the second half of 2014. This slowdown is in line with the fact stated in the introduction that growth in smartphone sales is expected to slow down and fall below double-digit percentage terms.

The estimated smooth curve corresponds to the first column in Table 1. Asymptotically,

---

[9]https://en.wikipedia.org/wiki/List_of_Wikipedias

Figure 1: Data and Estimated S-Curve for English Language Visits

Ratio of mobile to all visits of English Wikimedia pages over time
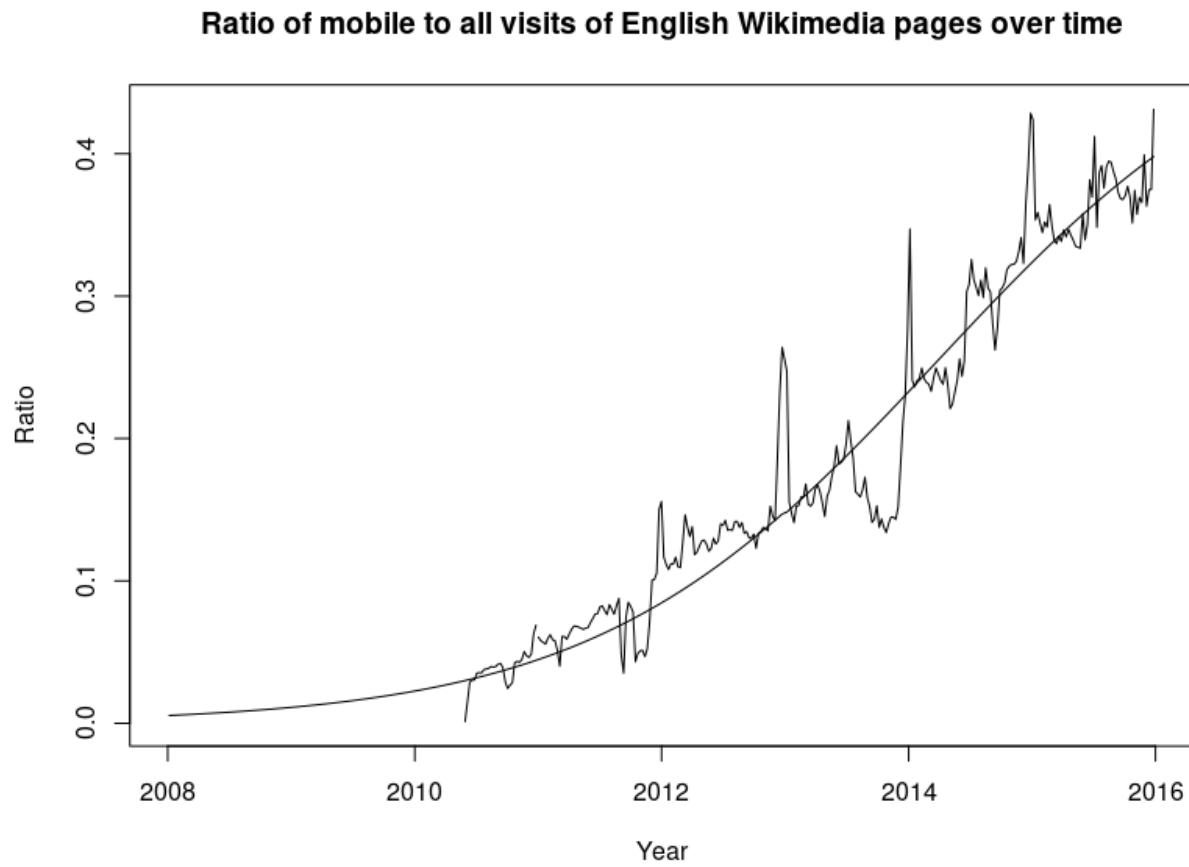
mobile traffic is estimated to make up just over 50% off all English language Wikimedia traffic. α determines the horizontal position of the S-curve and does not have a direct interpretation besides its possible use in calculating the point at which the rate exceeded 10% for the first time, which Griliches interprets as the effective introduction date of a technology. Finally, β measures the rate of technological adoption.

Table 1: Select Results for estimated S-curve parameters

	English	Swedish	German	Dutch	French	Russian	Italian
K	0.5094***	0.5032***	0.4501***	0.5383***	0.5306***	0.5124***	0.6113***
	(0.0207)	(0.0250)	(0.0133)	(0.0287)	(0.0254)	(0.0295)	(0.0131)
α	−4.5281***	−4.4558***	−5.5034***	−5.2449***	−5.1373***	−5.1640***	−6.3791***
	(0.0926)	(0.1184)	(0.1175)	(0.1053)	(0.0952)	(0.0849)	(0.1138)
β	0.0139***	0.0140***	0.0170***	0.0148***	0.0146***	0.0141***	0.0194***
	(0.0005)	(0.0007)	(0.0005)	(0.0006)	(0.0005)	(0.0005)	(0.0005)
R^2	0.9314	0.8916	0.9586	0.9441	0.9501	0.9519	0.978

Note: *p<0.1; **p<0.05; ***p<0.01

Referring to the other columns of Table 1, these parameters are roughly in line with those of other languages. For the top languages, the ratio of mobile to total Wikimedia activity is estimated to converge to between 45% in the case of German articles and 61% in the case of Italian articles. The estimated rates of technological adoption, β are similarly close to each other and range from 0.0139 to 0.0194. For all languages in Table 1, the estimated rates are highly significant and the S-curve model describes observed variations in relative mobile internet activity very well as the values for R^2 suggest. The consistency of the results from this baseline S-curve provide confidence for further exploration of the data based on this model.

In conclusion, Table 1 shows the estimated parameters for various languages. The languages chosen here represent the top ten languages in terms of the number of Wikipedia articles that are written in them. The numerical least squares minimization did not converge for Cebuano, Spanish and Waray.[10] These languages are excluded from Table 1.

---

[10]Cebuano and Waray are both languages native to the Philippines.

5.2 The Effect of Different Generations of Mobile Devices

The main goal of this paper is to identify the effect that mobile devices have on mobile internet usage. Specifically, introduction dates of new Apple iPhones and new Samsung Galaxy S phones (as a proxy for the Android segment of the mobile market) are used for the creation of dummy variables. For example in Table 2, the coefficient on iPhone 3GS captures the estimated effect of being in a week after the release date of the iPhone 3GS and before the release date of its successor, the iPhone 4. In terms of actual dates, the dummy on iPhone 3GS is equal to one if a week falls between June 19, 2009[11] and June 24, 2010 and equal to zero otherwise. Therefore, there is no overlap within the iPhone dummies or within the Galaxy S dummies but there is overlap between the dummies for Apple and Samsung phones since they belong to different product families. Note that the first two generations of the iPhone are not included in this analysis because they were introduced before Wikimedia started recording mobile traffic separately from regular page visits.

After including the dummies for the product cycles of these two manufacturers, the estimates of the basic parameters of the S-curve change quite drastically. The biggest change can be observed in the estimate of the parameter K, the ratio of mobile to total visits that internet traffic converges to. The predicted ratios for Swedish and Italian, in particular, are much larger than those in Table 1. However, the last in-sample observation is that of the last week of 2015. During that week, the iPhone 6S and the Galaxy S6 were the most recent generations in their respective product families and the dummies for these devices would have been equal to one, meaning the effect of the device has to be added to the S-curve. Using the effect of these two devices as a best guess for the effect of future devices and adding them to K, the sum is much closer to the estimates of K in Table 1.

This suggest that some caution has to be exercised when interpreting the results in Table 2. K does not correctly capture the asymptotic ratio anymore. Since the absence of any device effect (all dummies being equal to zero) represents the average effect of all generations of a device on mobile internet usage[12], K now represents the asymptotic ratio that would have been reached had this average effect persisted over time. Normally we assume diminishing returns in the effect of new technology over time. This can be verified

---

[11]http://www.apple.com/pr/library/2009/06/08Apple-Announces-the-New-iPhone-3GS[...].html

[12]As with any dummy variable model, either one dummy is left out and therefore captured by the intercept, or a full set of dummies is included and the intercept is dropped. In this case, the case where all dummies are equal to zero represents the average across all dummies. The model in this paper does not have an "intercept" and therefore the case in which all dummies are equal to zero (a week that does not lie within two release dates) captures the average and the effects of the dummies are consequently centred around zero.

Table 2: Select Results for estimated S-curve parameters and technology effects

	English	Swedish	German	Dutch	French	Russian	Italian
K	0.6202***	0.8508***	0.5293***	0.7385***	0.6756***	0.5020***	0.8329***
	(0.0639)	(0.1236)	(0.0592)	(0.1028)	(0.0686)	(0.0530)	(0.0591)
α	−6.5591***	−6.2031***	−6.1319***	−6.4940***	−6.6076***	−7.1199***	−7.7842***
	(0.6837)	(0.6011)	(0.6169)	(0.6511)	(0.6170)	(0.7955)	(0.6374)
β	0.0203***	0.0173***	0.0189***	0.0180***	0.0194***	0.0213***	0.0234***
	(0.0023)	(0.0022)	(0.0022)	(0.0023)	(0.0021)	(0.0026)	(0.0020)
iPhone 3GS	−0.0071	−0.0100*	−0.0079*	−0.0068	−0.0066	−0.0035	−0.0039
	(0.0046)	(0.0060)	(0.0046)	(0.0046)	(0.0041)	(0.0031)	(0.0032)
iPhone 4	0.0089	0.0054	−0.0049	0.0013	0.0005	−0.0019	−0.0010
	(0.0140)	(0.0169)	(0.0129)	(0.0135)	(0.0122)	(0.0101)	(0.0117)
iPhone 4S	0.0169	0.0278	0.0075	0.0153	0.0088	0.0081	0.0138
	(0.0173)	(0.0202)	(0.0158)	(0.0163)	(0.0150)	(0.0126)	(0.0146)
iPhone 5	−0.0040	0.0105	0.0065	0.0044	−0.0046	0.0223	−0.0011
	(0.0210)	(0.0242)	(0.0190)	(0.0196)	(0.0182)	(0.0155)	(0.0184)
iPhone 5S	−0.0454*	−0.0798***	−0.0234	−0.0465**	−0.0600***	0.0211	−0.0485**
	(0.0241)	(0.0286)	(0.0220)	(0.0229)	(0.0210)	(0.0178)	(0.0210)
iPhone 6	−0.0530**	−0.0857***	−0.0260	−0.0605**	−0.0876***	−0.0031	−0.0856***
	(0.0263)	(0.0319)	(0.0242)	(0.0254)	(0.0230)	(0.0193)	(0.0228)
iPhone 6S	−0.0787***	−0.1587***	−0.0380	−0.1063***	−0.1327***	−0.0253	−0.1299***
	(0.0291)	(0.0349)	(0.0265)	(0.0281)	(0.0255)	(0.0216)	(0.0257)
Galaxy S1	0.0202	0.0109	0.0066	0.0061	0.0128	0.0165*	0.0071
	(0.0124)	(0.0150)	(0.0114)	(0.0120)	(0.0109)	(0.0091)	(0.0109)
Galaxy S2	0.0285*	0.0151	0.0027	0.0049	0.0088	0.0168	0.0056
	(0.0147)	(0.0173)	(0.0133)	(0.0139)	(0.0127)	(0.0108)	(0.0126)
Galaxy S3	0.0241	0.0400*	−0.0033	0.0101	0.0083	0.0065	0.0046
	(0.0182)	(0.0208)	(0.0163)	(0.0168)	(0.0156)	(0.0135)	(0.0160)
Galaxy S4	−0.0249	0.0142	−0.0253	−0.0092	−0.0212	−0.0539***	−0.0327*
	(0.0220)	(0.0253)	(0.0200)	(0.0204)	(0.0190)	(0.0164)	(0.0197)
Galaxy S5	−0.0358	−0.0164	−0.0242	−0.0176	−0.0286	−0.0804***	−0.0623***
	(0.0250)	(0.0300)	(0.0230)	(0.0240)	(0.0219)	(0.0184)	(0.0219)
Galaxy S6	−0.0753***	−0.0643*	−0.0447*	−0.0592**	−0.0504**	−0.0714***	−0.1032***
	(0.0275)	(0.0329)	(0.0251)	(0.0263)	(0.0240)	(0.0203)	(0.0243)
R^2	0.9573	0.9415	0.9653	0.9598	0.9659	0.9685	0.9839

Note: *p<0.1; **p<0.05; ***p<0.01

here as well: The effect of the first few generations of iPhones and Android phones generally deviates slightly positively from the average effect and then drops off to negative deviations that are significantly different from the average effect in the last two or three generations across all languages. Therefore, the asymptotic ratio as time progresses is now represented by K plus at most the sum of the effects of the latest iOS and Android devices, which, as argued, is reasonably close to the estimates in Table 1.

The question that remains is whether the average effect of mobile devices on mobile internet usage is positive or negative. First, it should be noted that the time frame is still the same as in the model without dummy variables. Second, across all languages the estimate of α is smaller in the second model, meaning that the centre (the point of symmetry) of the logistic function is shifted to the right. This means that without the introduction of mobile devices the "natural" adoption of technology into mobile internet usage started later than previously estimated. Third, even higher levels of K are estimated. Given the same time frame as before, this requires the fourth piece of evidence: higher rates of technological adoption, which can be seen in the larger estimates of β. Combining these four facts results in a strong argument in favour of a positive average effect. Were the opposite true for any of these four arguments, the case would be much less compelling. Instead, if the average effect is positive, a higher β is expected since it captures the effect of time and the average effect of the various generations of a device. Likewise, a larger K corresponds to the asymptotic ratio should the average (positive) effect of mobile devices persist. Lastly, if the effect of mobile devices is positive and diminishing, it is the largest in early generations. That means that a fair share of early mobile traffic is explained by these devices. Consequently, the remaining ratio to be explained by the S-curve is much smaller early on in the time series, resulting in a shift to the right of the curve which is captured by the decrease in α.

In conclusion, there is ample evidence that the average effect of the introduction of mobile devices (both iPhones and Android phones) on mobile internet usage is positive but decreasing over generations.

Tables 3 and 4 in the appendix provide estimates for models that only include iPhones or Galaxy S devices respectively. These models were intended to be used as a robustness check for the results presented in Table 2. However, the model becomes highly unstable under those specifications and the results significantly deviate from estimates in both Table 1 and Table 2. This is a problem inherent to the release schedule that Apple and Samsung abide by. Since there is no overlap between the dummies for mobile phones within a company, and since each company releases their flagship phone almost precisely one year after its

predecessor, the dummies (almost) perfectly correspond to yearly cycles. This would not be a problem in a linear model as partial effects are unaffected by other variables. However, in the nonlinear model underlying this paper, the growth rate (the partial effect of time) is not constant. Therefore, if the growth rate has any patterns that correspond to annual cycles (such as holidays etc.), these can be incorrectly captured by the device introduction dummies.

This problem is avoided in the model that includes Apple as well as Samsung devices since there is some overlap between the dummies due to the fact that the two companies operate on different release schedules. This suggests that, ideally, one would include many different device families in the analysis in order to have more overlap between the dummy variables. The problem with this approach is that the resulting loss of degrees of freedom has quite significant consequences for the numerical minimization of the sum of squared residuals in the sense that the approximations are much less likely to converge.

6   Conclusion

When describing the ratio of mobile Wikimedia page visits relative to the total number of visits as a proxy for overall mobile internet usage, the S-curve model for technology diffusion first used by Griliches (1957)[1] performs remarkably well. I find that, for the most popular languages on Wikipedia, this ratio is predicted to converge to somewhere in the 50% to 60% range depending on the language. Furthermore, I find rates of adoption that are similar between languages and typically lie between 0.015 and 0.02.

In addition to this model purely described by the progression of time, I find evidence that strongly suggests that the introduction of iPhones and Samsung Galaxy S phones (as representative for the Android segment of the mobile market) have a generally positive effect on the relative usage of mobile internet. More specifically, I find that this effect is diminishing over time, especially with the last two generations of phones. The effect on the ratio from the last two generations is significantly less than the average effect across generations and it is not clear if the absolute effect is positive or negative. It is, in fact, possible that the negative deviation from the average effect is greater in magnitude than that average so that the overall effect of the latest generations might actually be negative. While this may seem counter-intuitive at first, it may be explained by increased usage of specialized apps. The Wikimedia page traffic data dumps underlying this paper only track browser traffic and do not include traffic resulting from the official and/or unofficial Wikipedia apps. However,

this does not have to be a flaw or caveat in the analysis since it would probably not be unreasonable to believe that a shift to apps away from mobile browsers is also representative for general mobile usage and not unique to Wikipedia.

Finally, it would be desirable to have some better robustness checks. Two approaches come to mind that may achieve this. First, as mentioned above, using only one of the manufacturers should work but fails due to very regular release schedules. The closest substitute to this approach is using different manufacturers and device families. However, given the large market share that Apple has, ignoring iPhones does not seem prudent. Alternatively, one may use a different manufacturer as a proxy for all Android devices. The problem with this approach is that there is no single family of devices that could serve as a reliable proxy. For example, the Sony Xperia series, though in existence for many years, had very inconsistent sales figures over its generations[13] and the Nexus family only moved away from the image of a developer phone to a phone for the general public with the Nexus 4 at the end of 2012.

A second approach for a robustness check could be to use the amount of data transmitted to mobile and non-mobile devices to calculate a ratio similar to the ratio of visits used in this paper. This method has the advantage that the data is also published in the page traffic dumps. However, there is about a 90% permanent drop in the calculated data ratios for all languages in August 2011. Considering that the ratio of visits, calculated in the same way from the same data source, does not experience this jump, it is likely that something changed in the way data transmission was measured by the Wikimedia foundation. While it would be relatively easy to correct this structural break in the data, I did not feel comfortable doing so without knowing the reason for the sudden drop in the ratio. Unfortunately, I was not able to find out what caused this and therefore decided against using the data even though I fully recorded it and it is available together with the other data.

---

[13] "Due to certain changes in Sonys organizational structure, sales and operating revenue and operating income (loss) of the MC segment of the comparable prior period have been reclassified to conform to the current presentation." as an example from the Sony Corporation Consolidated Financial Results for the Third Quarter Ended December 31, 2015

References

[1] Griliches, Zvi, 1957, "Hybrid Corn: An Exploration in the Economics of Technological Change", *Econometrica*, Vol. 25-4, pp. 501–522.

[2] Caselli, Francesco, and Wilbur Coleman II, 2001, "Cross-country Technology Diffusion: The Case of Computers", *American Economic Review* 92-2, pp. 328–335.

[3] Christensen, Clayton M., 1992, "Exploring the Limits of the Technology S-Curve. Part I: Component Technologies", *Product and Operations Management* 1-4, pp. 334–357.

[4] Czernich, N., Falck, O., Kretschmer, T., and Woessmann, L., 2011, "Broadband Infrastructure and Economic Growth", *Economic Journal* 121, pp. 505–532.

[5] Majumdar, Sumit, and Venkataraman S., 1992, "Network Effects and the Adoption of New Technology: Evidence from the U.S. Telecommunications Industry", *Strategic Management Journal* 19, pp. 1045–1062.

[6] Hall, Bronwyn, and Khan, Beethika, 2002, "Adoption of New Technology", *New Economy Handbook*.

[7] Wikimedia Foundation, "Page view statistics for Wikimedia projects", `https://dumps.wikimedia.org/other/pagecounts-raw/`, accessed June–July 2016.

Appendix

Table 3: Select Results for estimated S-curve parameters and technology effects of iPhones

	English	Swedish	German	Dutch	French	Russian	Italian
K	0.2792***	0.6026***	0.4192***	0.5246***	0.5357***	0.2497***	0.6090***
	(0.0222)	(0.0553)	(0.0339)	(0.0469)	(0.0410)	(0.0176)	(0.0314)
α	−15.0766***	−5.3958***	−6.6824***	−6.4512***	−7.0195***	−19.2966***	−8.1663***
	(1.4357)	(0.4171)	(0.6301)	(0.5838)	(0.5836)	(2.6636)	(0.5272)
β	0.0462***	0.0162***	0.0208***	0.0190***	0.0206***	0.0540***	0.0248***
	(0.0043)	(0.0015)	(0.0021)	(0.0020)	(0.0019)	(0.0077)	(0.0017)
iPhone 3GS	0.0014	−0.0136**	−0.0039	−0.0052	−0.0031	0.0012	−0.0018
	(0.0031)	(0.0059)	(0.0034)	(0.0039)	(0.0032)	(0.0023)	(0.0028)
iPhone 4	0.0567***	0.0092	0.0097	0.0097	0.0198***	0.0286***	0.0119***
	(0.0027)	(0.0093)	(0.0059)	(0.0066)	(0.0052)	(0.0020)	(0.0039)
iPhone 4S	0.1139***	0.0381***	0.0282**	0.0272**	0.0362***	0.0648***	0.0382***
	(0.0035)	(0.0147)	(0.0111)	(0.0118)	(0.0100)	(0.0024)	(0.0084)
iPhone 5	0.1398***	0.0311	0.0274*	0.0171	0.0274*	0.1079***	0.0332**
	(0.0078)	(0.0196)	(0.0155)	(0.0164)	(0.0147)	(0.0030)	(0.0141)
iPhone 5S	0.1040***	−0.0587**	0.0007	−0.0318	−0.0247	0.1364***	−0.0036
	(0.0155)	(0.0242)	(0.0181)	(0.0194)	(0.0175)	(0.0078)	(0.0172)
iPhone 6	0.1037***	−0.0403	0.0082	−0.0272	−0.0397**	0.0971***	−0.0186
	(0.0181)	(0.0271)	(0.0199)	(0.0213)	(0.0191)	(0.0107)	(0.0188)
iPhone 6S	0.1014***	−0.0923***	−0.0026	−0.0595**	−0.0807***	0.0852***	−0.0536**
	(0.0214)	(0.0299)	(0.0227)	(0.0242)	(0.0219)	(0.0138)	(0.0218)
R^2	0.9497	0.9363	0.9635	0.957	0.9635	0.9624	0.9819

Note: *p<0.1; **p<0.05; ***p<0.01

Table 4: Select Results for estimated S-curve parameters and technology effects of select Android phones

	English	Swedish	German	Dutch	Russian	Italian
K	0.4755***	0.5060***	0.4452***	0.2895***	0.4224***	0.6232***
	(0.0384)	(0.0799)	(0.0329)	(0.0281)	(0.0226)	(0.0341)
α	−6.6044***	−6.4847***	−6.1796***	−14.9390***	−6.9469***	−7.2488***
	(0.5217)	(0.7432)	(0.4111)	(1.7066)	(0.4176)	(0.4404)
β	0.0201***	0.0181***	0.0191***	0.0427***	0.0227***	0.0223***
	(0.0018)	(0.0026)	(0.0015)	(0.0050)	(0.0014)	(0.0015)
Galaxy S	0.0334***	0.0266***	0.0049	0.0237***	0.0123***	0.0052
	(0.0052)	(0.0070)	(0.0048)	(0.0032)	(0.0037)	(0.0044)
Galaxy S2	0.0570***	0.0592***	0.0134*	0.0556***	0.0124*	0.0141*
	(0.0086)	(0.0110)	(0.0077)	(0.0030)	(0.0065)	(0.0077)
Galaxy S3	0.0630***	0.1094***	0.0214*	0.1081***	0.0022	0.0193
	(0.0137)	(0.0171)	(0.0120)	(0.0041)	(0.0106)	(0.0130)
Galaxy S4	0.0151	0.0697***	−0.0036	0.1204***	−0.0637***	−0.0165
	(0.0182)	(0.0234)	(0.0163)	(0.0097)	(0.0136)	(0.0171)
Galaxy S5	0.0192	0.0624**	0.0022	0.1188***	−0.0854***	−0.0202
	(0.0213)	(0.0287)	(0.0193)	(0.0160)	(0.0154)	(0.0195)
Galaxy S6	−0.0110	0.0454	−0.0097	0.0960***	−0.0597***	−0.0396*
	(0.0236)	(0.0318)	(0.0213)	(0.0194)	(0.0172)	(0.0219)
R^2	0.9521	0.9155	0.962	0.9505	0.9665	0.9805

Note: *p<0.1; **p<0.05; ***p<0.01

When only using Galaxy S phones, the numerical optimization did not converge for visits of French Wikimedia pages.