DRAFT: Not for Citation

Measuring the Production of User-Generated Internet Content

By

Jon Samuels and Rachel Soloveichik[1]

December 2017

Introduction

"Free" websites, such as Facebook, Reddit, YouTube and Wikipedia combine three separate categories of "free" digital content: 1) platforms for storing content and distributing it to users (the underlying software and cloud computing), 2) professionally-generated content, such as articles written by advertisers or marketers, and 3) user-generated content, which is produced by amateurs as a hobby without any expectation of payment. User-generated content includes Facebook comments as well as comments hosted on newspaper sites and blogs, Reddit posts, (some) YouTube videos and Wikipedia articles. Our previous research on "free" content (Nakamura, Samuels, & Soloveichik, 2017) introduced an experimental GDP methodology which includes "free" platforms and "free" professionally-generated content in measured GDP. The paper then recalculated nominal GDP, real GDP and total factor productivity using the experimental methodology. Between 2005 and 2015, our experimental GDP growth is 1.5% per year rather than the official growth rate of 1.4%.

User-generated content is produced outside the market sector, so our previous research intentionally excluded it from measured GDP. This paper does not attempt to revisit the current distinctions between official GDP (constructed to be consistent with the System of National Accounts) that reflects market transactions, home production, and leisure. According to (Moulton, 2015), household own-account non-housing services production (e.g., vacuuming, child care, home meal preparation) is excluded from the SNA because most services are self-contained activities, typically there are no suitable market prices for valuation, and production of this sort generally does not influence

[1] Jon Samuels and Rachel Soloveichik are research economists at the U.S. Bureau of Economic Analysis; jon.samuels@bea.gov and rachel.soloveichik@bea.gov. The views expressed in this paper are those of the authors and do not represent those of the U.S. Bureau of Economic Analysis, or the U.S. Department of Commerce. We thank Ana Aizcorbe for her help in obtaining the TUP data and useful discussions, Dan Ness of Metadata for providing and explaining the TUP data, and Dylan Rassier for useful comments.

economic policy. Nevertheless, there is a long-standing general interest in satellite accounts that are broader in scope than the official GDP statistics (Abraham & Mackie, 2005), (Bridgman et. al., 2012), (Bridgman, 2016). In this paper, we take the current distinctions as given and analyze user-generated content in the framework of household production. The ultimate purpose of this line of research is to create an experimental account tracking nominal inputs, real inputs, nominal output and real output of user-generated content. The objective of current paper, however, is to present some exploratory statistics on the labor, capital and value-added of user-generated content.

We find that the number of American adults creating content quadrupled from 43 million in 2006 to 166 million in 2016. Both sexes and all age groups experienced a large increase in content creation, but the growth rate is especially large for older adults because they started from a much lower base in 2006. Across the entire population, we estimate that content creation time quintupled from 45 hours per person in 2006 to 251 hours per person in 2016. In comparison, the American Time Use Study (ATUS) shows that the work time fell slightly over the same time period. In particular, market sector work fell from 1,369 hours per person in 2006 to 1,318 hours per person in 2016 and household production work fell from 1,219 hours per person in 2006 to 1,208 hours per person in 2016.

This paper will be divided into 4 parts. Section 1 will provide background on "free" digital media and user-generated content. Section 2 will briefly describe the data used in this paper. Section 3 will document preliminary results on labor time and capital inputs devoted to creating user-generated content. Section 4 will briefly discuss the pre-existing household satellite accounts and compare our results on user-generated content. Finally, we will conclude with a review and plan for future research.

1. Background on "Free" Digital Content and National Accounting

1.1 "Free" platforms and "free" professionally-generated content

To frame ideas, it is useful to revisit our earlier work on the production of "free" digital content in (Nakamura, Samuels, & Soloveichik, 2017). In that paper, we analyze how "free" platforms and "free" professionally-generated content, which are supported by either advertising or marketing, relate to the GDP accounts. To be clear, our earlier research covers a wide range of "free" digital content including Google search, NBA.com, downloadable apps and more. Our basic argument is that the current treatment in the GDP and productivity accounts does not capture the value of this content because consumers and businesses do not pay directly for the content. To circumvent this, we modify the GDP and productivity accounts to recognize the implicit barter transaction that occurs between the providers of the digital content and the users of the content: viewers receive "free" content in exchange for advertising (in the case of Google search) and for marketing (in the case of NBA.com). We then value this content at the cost of production, which is the standard method for measuring production value when outputs are not directly measured.

The jumping off point of this paper is that the barter transaction model discussed above covers only production which occurs within the market sector. Only that production is considered in scope for the GDP accounts and only that production is ultimately counted in personal consumption expenditure, investment, government spending[2], or net exports (C+I+G+X-M). Take Twitter for example, as a producer of "free" platform services whose only source of revenue is advertising. In BEA's input-output accounts, this advertising revenue is the nominal value of Twitter's output. On the input side of Twitter are the intermediate expenses such as materials and energy, and value added. In the framework of the production account, value added encompasses payments to factor services, capital and labor. These are the costs of production, and this is what the barter transaction in (Nakamura, Samuels, & Soloveichik, 2017) covers.

User-generated content is different from the "free" platforms and "free" professionally-generated content studied in our earlier research. Professionally-generated content that is produced by private sector businesses is done so with the expectation that this content will yield sufficient revenue to cover production costs. As a result, we treat these content types as part of the market sector and include them in measured GDP. In contrast, user-generated content is produced without any expectation of revenue or other material reward.[3] Conceptually, content creation is a type of volunteer activity which is intended to benefit a community, the environment or another social purpose. Like other volunteer activities, this output is not part of the market sector and therefore not counted in official GDP.

In this setup, the value of the content that is posted by Twitter users is not part of Twitter's production costs. While it may tempting to think that the value of amateur Tweets is reflected in Twitter's advertising revenue, the actual Tweets are not produced by the business sector but instead are user-generated content. Conceptually, we assume a competitive environment with multiple possible "free" platforms, minimal network effects and minimal interactions between user-generated content and the "free" platform software. In that environment, users gravitate towards the "free" platform software which provides the best content hosting features in return for their ad viewership. In that simplified model, Twitter's advertising revenue is only sufficient to cover its platform management costs. Furthermore, user-generated content does not create organizational capital or other long-term value to the "free" platforms that host it.[4] In other words, the problem of valuing "free" digital content which is supported by advertising and marketing is separable from the problem of valuing user-generated content.

[2] "Free" digital content created by either governments or non-profits is **already** counted in measured GDP based on its production costs. Accordingly, our previous paper focused on content created in the private business sector.
[3] The line between professionally-generated and user-generated content is sometimes fuzzy. For example, a writer might start a blog as a hobby, and then start earning money from recommendations or selling articles. However, the majority of hobbyists remain amateurs. For now, we treat all user-generated content as amateur.
[4] In a more complex model, there may be network effects and other factors which allow a few lucky companies to earn extraordinary profits. However, these extraordinary profits may be offset by the losses from startups which are temporarily offering low cost services to build market share and the unlucky companies which fail completely.

1.2 Constructing accounts for user-generated content

In this section, we explore the role of user-generated digital content in relation to the official GDP accounts and the BEA satellite on household production. The basic premise of the BEA satellite account on household production is that much of household work has a clear counterpart in the market sector and thus shifts in preferences between work within the home and outsourced home services affect measured GDP. To address this, (Bridgman, 2016) provides a modified GDP account that includes production of nonmarket services within the home. That satellite account generally follows the classifications in the American Time Use Survey (ATUS) when allocating activities between official GDP, household production, and leisure. The ATUS currently classifies user-generated content as a leisure activity and therefore it is excluded from measured household production.

An important conceptual issue is that these imputations should appear on both the output and input side of the domestic production account. Imputed labor services are combined with imputed capital services from consumer durables to create new consumption goods. This output value is over and above the value of the purchased intermediates that are already in the GDP accounts. Accordingly, cooked meals are assumed to have a value over and above the value of purchased foods that are already in the GDP accounts. That is, the imputed labor services and imputed capital services account for the value added of domestic household production.

This paper experiments with a similar approach for user-generated digital content. On the input side, producers of user-generated content employ (their own) labor services and a portion of the capital services of consumer durable services to produce content. We measure the labor services component using data from the TUP survey that will be discussed later in the paper. Of course, a simple count of hours does not capture the wide quality range of user-generated content. A drunken Facebook comment likely has much less value than a carefully constructed product review on Amazon.com. It is also likely that output quality may vary randomly. For example, a joke picture might go viral or it might fall flat. Nevertheless, we start with the approach that user time is a reasonable proxy to impute the labor services used to create user-generated content.

Content creators often share their output on specialized websites devoted to user-generated content. Each specialized website has its own norms governing membership policies, the format of content and the tone. For example, a Facebook discussion group may be restricted to invited members and supervised by volunteer moderators who delete posts which break the discussion group rules. In many cases, private companies like Amazon maintain websites and set norms for their own purposes. There is a rich economic literature studying these norms and modeling the advantages and disadvantages of each norm. Most recently, the paper “Fast and Slow Learning from Reviews” (Acemoglu, et. al. 2017) provides a useful overview of that literature. Our paper is focused on measuring labor and capital inputs to user-generated content, not determining which website norms are best.

1.3 What's excluded from our experimental account?

We do not study business sector user-generated content because the value of that content is already included in the official GDP statistics and the time spent creating the content is already included in official labor input. For example, most of the participants in the open source software community are professional programmers who are sharing on-the-job content. In theory, open source software could be modeled as a barter transaction: programmers provide software in return for software. However, such a barter transaction would raise measured output and measured intermediate expenses by the exact same amount with no net effect on value-added or final output.[5] As a result, tracking business user-generated content would complicate BEA's existing input-output tables without changing final output at all. For simplicity, we will not include it in this paper.

We also exclude amateur user-generated content that is used by the business sector. For example, amateur kayakers designed and tested new models that freely shared with commercial manufacturers (von Hippel 2017). Conceptually, this type of transaction could be modeled as a gift from the household sector to the business sector.[6] However, neither the official GDP statistics nor the existing satellite accounts are set up to handle the business sector receiving volunteer services – and we have not yet been able to develop a consistent accounting methodology to handle such volunteer services. Accordingly, we will not consider this transaction in this paper draft. Instead, we will assume that amateur user-generated content is used by the household sector.

Finally, our experimental account does not track data exhaust. By definition, data exhaust is automatically produced by ordinary business transactions. For example, retailers use computerized cash registers to ring up purchases and accept payment. These transactions generate terabytes of data on the exact items purchased, returns, and other useful information. Many retailers use this information to price existing products, change story layouts or even design new products. Economics researchers also use data exhaust to track the overall economy and answer research questions. Regardless of which companies produce data exhaust, how much they spend collecting and organizing the data, and how it is used, data exhaust requires neither user time nor user capital inputs to create. As a result, it is assigned zero value in our experimental account tracking user-generated content.[7]

Note that our estimates do not measure social welfare. In general, economic theory allows goods and services to yield consumer surplus far beyond their production costs. Furthermore, volunteer activities

[5] By construction, the net nominal value of barter transactions is always zero. It is possible that individual industries may have different price indexes for the user-generated content they create and the user-generated content they use, so there is an impact on real value added and total factor productivity for that industry. But the net impact on aggregate real value-added and total factor productivity is likely to be small.

[6] The gift model only applies if the business sector provides absolutely nothing in return for the amateur user-generated content. Many businesses offer "free" platforms for users to discuss their products, and other valuable content to the hobbyist community in return for their designs. This type of barter transaction is covered in our earlier paper on advertising and marketing supported content (Nakamura, Samuels and Soloveichik 2017).

[7] Data exhaust may represent intangible capital for the companies which own it. The System of National Accounts 2008 does allow countries to capitalize databases (Section 10.112-10.114), but measuring the value of such databases is extremely difficult. BEA has not yet implemented SNA's recommendation.

like user-generated content are especially hard to value. Some economic models predict that volunteering helps improve overall social functioning and thereby provides social benefits far exceeding its private cost. Other economic models predict that the social benefit of volunteering is nearly zero because the supposed beneficiaries don't actually want the services provided. User-generated content is often created to bolster one side in a controversy over partisan politics, cultural expression, romantic parings for fictional characters and other zero sum arguments. Furthermore, user-generated content may sometimes create negative externalities like increased social atomization. Our paper will not attempt to measure the welfare impact of either the production or consumption of user-generated content. Instead, we will simply measure the time and capital devoted to its production.

2. Description of the Dataset Used

For this exploratory work, we rely on data from the Technology User Profile, a proprietary dataset available for purchase from Metafacts. The major advantage of this survey is that it asks questions on the activities people use their devices for; this allows us to identify activities that are associated with user-generated content. The TUP has survey data available back to 2001 (though it began in 1983), but we begin our analysis in 2006 because this is the first year that we have access to in which the survey contains information on hours spent using each device. As a point of reference this first IPhone was released in 2007.

The TUP is a representative sample of adults that own connected devices and includes weights that are constructed to yield totals for the people in the United States. The survey is based on two phases: the first phases builds a sample frame and develops estimates of technology usage (PC, Cell, Tablet, Gaming Device) by demographic groups, in particular gender and age groups and then Metafacts implements the sample within each demographic group based on a total target sample. Beginning in 2006, the initial screening was conducted by telephone on a nationally representative set of adults, and the subsequent survey was conducted online. As in example, in 2016, the total target sample was 7500 completed surveys and age group 35-44 had a target of 1,207 completed surveys. Metafacts adjusts the weights to be representative of the target population by compensating for varying response rates and reweighting to be consistent with the entire population of U.S. adults.

The survey asks detailed questions about the respondent and the respondent's use of technology and technology devices. Once screened in, the survey asks people about the websites they have visited, their computer literacy, employment status, internet access, and a series of questions on which devices the respondent uses, which activities are performed on the devices, and how much time is spent on each device. Our focus is on the hours spent using each device and the activities performed on each device. We discuss below how we use this data to measure the production costs associated with user-generated content.

The TUP survey is not comparable to the ATUS. The TUP is focused on tracking time usage by device, and so it includes work time, household production time and leisure time. For example, a self-employed

individual might use their computer to contact clients during the day, look up recipes in the afternoon and then play games at night. In contrast, the ATUS allocates computer usage for work into the broader code for work, computer usage for cooking into the broader code for cooking, etc. In addition, the ATUS questions appear to focus on computers and may miss time spent with digital devices like video game systems or smartphones. Finally, the ATUS focuses on primary activities and frequently misses secondary device usage. For example, the ATUS might report an individual spent 30 minutes on a primary activity of eating dinner with family – but completely miss time spent checking Google to settle arguments. Because the ATUS has such a different focus, we will not use its data on computer time at all.

The TUP survey is also very different from the Forrester data used in our earlier paper on "free" platforms and "free" professionally-generated content (Nakamura, Samuels and Soloveichik 2017). That paper was focused on media consumption, so we only purchased the Forrester questions tracking media time usage. Those Forrester questions split personal Internet time from work Internet time – but they do not track offline device usage. We do not know if Forrester's other survey questions match the TUP data. Furthermore, the line between using a device online, using a device offline and not using a device are very fuzzy. For example, a commuter might use Waze to pick the fastest route home. Are they considered online for their entire drive – or just online when they check the route? Are individuals using a smartphone when it's in their pocket waiting for calls? Survey respondents often use their own judgment when answer questions, and so there is no precisely "correct" answer.

3. Empirical Results

3.1 The extensive margin: people generating content and their demographics

Our first step in measuring the production of user-generated content is to estimate the number of people engaged in content production. We do this by tabulating the number of people involved in activities tied to user-generated content. Another issue that we encounter is that the activities included in the TUP change over time. An implicit underlying assumption that we make is that the TUP covers all relevant activities in a given year – so any omitted activities can be set equal to zero. For example, microblogging (e.g. Twitter) was not big enough to include in the activities until it became prominent enough for TUP to cover. The list of activities and those that we consider as content generating activities is available in Appendix Table A for 2016; other years are available upon request. User-generated content spans many different types of activities, some simple as in "liking" someone's post and some potentially more sophisticated like writing an individual blog or producing one's own personal webpage.

The TUP data also handles multiple devices in an inconsistent manner over time and across survey questions. Some of the TUP data series cover more than four devices used by the survey respondent, but this is not the case for all of the series. Thus, we only use the first four devices reported in the TUP in

2016.[8] In 2006, the survey covers only two personal computers, thus our time series estimates of the number of people generating content are subject to changes in the type and quantity of reported devices. Our take on this at this point is that the TUP reflects almost all of the devices and activities at a given point in time, thus we are hopeful that this coverage issue does not create a significant bias in our estimate. Nevertheless, we intend to attempt to address this in the future.

Our tabulations show that the number of people who were online and producing content grew from 43 million in 2006 to 166 million in 2016, a growth rate of 136%. It is possible that a few of the online content creators had previously been creating offline content. For example, some print newspaper readers might have written letters to the editor. However, new technologies like cloud computing, smartphones and social media software have made content creation and distribution much easier than it was before. As a result, many people have started actively creating content when they were previously passive consumers of professionally-generated content.

Figure 1 shows the proportion of people online and producing content. Among the online population, content creators grew from 28% of those online in 2006 to about 80% in 2016. This, to a first order, shows the tremendous growth in the production of user-generated content. This growth in reinforced by a 30% increase in the number of people online in the U.S., from 153 million people in 2006 to 208 million in 2016. This growth occurs for all demographic groups in the TUP.

Figures 2-4 show the proportion of the population online and producing content by age. In every year, young adults are more likely to be online and more likely to be creating content. But Internet usage and content creation have both spread across the generations. Most strikingly, older (55 and above) content creators grew from only 6% of the population in 2006 to 51% of the population in 2016.

Figure 5 shows the male share of the digital population. The main result is the males account for approximately half of the online population and half of the content creators. Early in the sample period, males were slightly more likely to be content producers. After 2006, there is a small decline in the share of content producers that were male between 2006 and 2013, perhaps as the ongoing recovery from the Great Recession disproportionally affected male workers, their jobs, and available time to spend producing online content. However, these fluctuations are relatively minor. In summary, men and women both experienced a dramatic in content creation over the past decade.

The results in Figures 2-5 suggest that content creation is distributed far more evenly than most household production. Traditional household production like cooking, cleaning and childcare is skewed toward women. Furthermore, older adults are much less likely to have young children – so they perform much less childcare. The results in Figure 4 indicate the older generation now produces a significant share of "free" content appearing online, implying that those interested in household production and its distribution should be cognizant of this role of older contributors. An open question is whether retirement and the production of online content are substitutes or complements.

[8] This covers PCS, Tablets, Cell Phones, Game Consoles, and Unknown, though no one in the 2016 survey reports producing content using Game Consoles or Unknown, according to the activities that we label as content production.

Figures 6 and 7 show an activity breakdown of user-generated content in 2010 and 2016. In both years, significantly more people engaged in simpler activities like sharing photos on their social network or posting comments on websites maintained by others. In addition, a large portion of the population posts product and restaurant reviews, providing important support for websites and apps like Yelp, Amazon, EBay, and Uber and other sites that provide indicators of reliability by having user sourced information. But complex activities like maintaining one's own website or sharing videos are also common and growing fast.

3.2 The intensive margin: time spent generating content

The foundation of our experimental production account is total hours spent on user-generated content. Due to limited data and as a first pass at assessing the potential magnitude of the production cost of user-generated content, we use a simple methodology. First, we divide all activities covered in the TUP into those that produce user-generated content and those that do not. For the years 2006, 2010, and 2016 we allocate time spend online (as measured by the TUP) into online time generating content and other time using the proportion of activities that generate "free" content. For example, if a survey respondent engaged in 30 activities online and 10 of them were those that are associated with the production of online content, then we would allocate 1/3 of that person's online time to the production of user-generated content.

Table 1 gives the average hours generating content by age group and gender over time. Across the subpopulation of content creators, user-generated content averaged 4.4 hours per week in 2006, 7.8 hours per week in 2010 and 6.9 hours per week in 2016. This intensive increase might appear small, but it occurred at the same time content creation numbers were skyrocketing. Across the entire population, user-generated content averaged 0.9 hours per week in 2006, 3.1 hours per week in 2010 and 4.8 hours per week in 2016.[9] At the aggregate, the ratio of hours spent generating content to economy wide hours worked increase from 3% in 2006 to 22% in 2016.

Over this period, average hourly time spent generating content was skewed towards the young. In 2006, Females (18-24) contributed the largest amount of time per hour on content generating activities (about 4 hours per week), while males (18-24) contributed the second largest amount of time per hour (about 3.5 hours per week). Males and females (24-34) were the next largest demographic groups in terms of time spent per week on content-producing activities. In 2016, younger females (18-24 & 25-34) exceeded the average amount of time spent by males of the same age group. Time spent on content generating activities increased for each gender and age group between 2006 and 2010 and 2010 and 2016. By 2016, our estimation method suggests that even the groups spending the least amount of time on activities that produce "free" content, males and females over 65, spent on average over two hours per week on content creation.

[9] To estimate total time, we impute time online for those that did not report this using the average for those that did report by sex and age group.

3.3 Capital inputs to user-generated content

Capital services play an important part in home production as well. (Bridgman, 2016) presents estimates of the service flow from consumer durables. In a full production account, this would appear both on the input side as a capital input, and on the output such as the consumption of capital services. Again, this is over and above the value of the purchased durables and represents the consumption of the capital services from the entire stock of consumer durables owned by the household. On the output side of the account, the services flow represents the consumption of these services, while on the input side, the imputation accounts for the (implicit) payments to capital to produce the services. These services flows are similar in nature to the service flow from owner occupied housing that is included in the official GDP accounts which can be viewed as a production process where the owner of the capital good rents out the good to the household for a period of time. In this framework, these services can be viewed as an equivalent rental payment to use the accumulated stock of household capital. A typical assumption is that the prices for the new output are the same as the prices for the inputs, so that there is no TFP measured in household production.

To measure the capital services used in the production of user-generated content, we start with the capital service flow from the stock of all household consumer durables. Because most capital goods used by the household are purchased, not rented, these flows need to be imputed. For our estimates, we rely on the dataset underlying the work of Jorgenson, Ho, Samuels, 2017 and Jorgenson, Nomura, Samuels, 2016. They impute the service flow by estimating the user cost of capital:

$$c_{h,i,t} = (r_{h,t} - \pi_{i,t} + (1+\delta_i)\pi_{i,t})P_{i,t-1}$$

where $c_{h,i,t}$ is the unobserved rental rate, $r_{h,t}$ is the rate of return in the household sector, $\pi_{i,t}$ is the asset specific capital gain, δ_i the asset's depreciation rate, and $P_{i,t-1}$ the lagged asset investment price. With a measure of $c_{h,i,t}$, the value of imputed capital services is simply the price $c_{h,i,t}$ times the net capital stock quantity of household consumer durables. [10] This yields a value of the service flow of consumer durables on the output and input side of the account. To get a sense of the magnitude, Jorgenson, Ho, and Samuels estimate a services flow from consumer durables of $1,150 billion in 2015 (compared to GDP of $18,924 billion).[11]

To estimate the value of capital input in the production of UGC, we start with the capital service flow from three types of assets: Computers, Communications equipment, and Software. In 2015, this was about $142 billion. We then split this into the portion used to generate UGC and other using the ratio of hours online spent generating UGC versus to total time online. Based on data in the TUP, the share of

[10] Jorgenson uses the rate of return from the private sector as a measure of the rate of return, while (Bridgman, 2016) uses the return on personal interest and dividend income.

[11] This is GDP from the integrated industry level production account which differs from GDP due to imputations of government capital services.

time online generating content increased from about 2.5% of device time in 2006 to 7.9% of device time in 2010 and to 9.4% of device time in 2016.[12] Using these shares, we estimate that capital services contributed about $2.0 billion in 2006, $6.8 billion in 2010, and $13.0 billion in 2015.[13] To reiterate, this imputation represents the addition to the value added of domestic production if the capital services devoted to user-generated content were included in the experimental account. Our experimental account does not yet track intermediate inputs devoted to user-generated content, so it does not measure gross output.

It is likely that real capital inputs have grown even faster than the nominal capital inputs calculated above. Smartphones barely existed in 2006 – but they are now nearly universal. In addition, there may be capital stocks of pre-existing user-generated content that can be repurposed. For example, old Tweets may appear initially to be jokes – but later provide valuable evidence of a behavior patterns.[14] These capital improvements have likely lowered the price of content creation. However, we have not yet found data measuring labor productivity for content creation. Accordingly, we are not able to estimate real production of user-generated capital for this paper draft.

Our approach to measuring user-generated content is based on time and capital spent on activities related to producing content. An alternative approach is to use available data on the actual quantity of user-generated content, such as the database of Amazon reviews maintained by Julian McAuley at http://cseweb.ucsd.edu/~jmcauley/. Because the breadth of user-generated content is difficult to assess, and because databases covering user-generated content are limited, we do not pursue this approach in this paper. In future versions of this paper, we plan to use Amazon reviews and search for other alternative data as a point of comparison to our measures based on production time. Of course, output based measures would permit an analysis of the productivity in producing user-generated content, a topic that we cannot address with our current methodology.

[12] This is based on the method described above of allocating total time online in proportion to activities generating content and other. Our capital dataset ends in 2015 and we interpolate the allocation shares between years of the TUP data.

[13] Our dataset for capital ends in 2015. We extend it to 2016 using the nominal stock of IT consumer durables as a proxy (BEA's Fixed Asset Table 8.1, line 12). The TUP focuses on Internet users, so it misses the offline population. We assume that these individuals rarely own computers, so we need not adjust the IT capital services downward.

[14] The discussion in this paper generally assumes that user-generated content is short-lived and so consumption tracks production closely. A more complex model would allow both positive and negative long-term value. For example, a Wikipedia article on a scientific topic might be useful for decades. Conversely, some user-generated content may retain negative value. For example, a photographic might seem romantic at the time but eventually prove to be horribly embarrassing. The net value of past content is theoretically ambiguous and will be set to zero.

4. Valuing Household Production and User-Generated Content

4.1 Existing research on household production

BEA's published GDP statistics exclude household services from measured output. This exclusion is consistent with the international guidelines for national accounts, the System of National Accounts 2008 (SNA 2008). However, BEA has previously estimated satellite accounts tracking traditional household services like cooking, cleaning and childcare (Bridgman 2016, Landefeld, Fraumeni and Vojtech 2009). This focus is consistent with the American Time Use Survey (ATUS), which provides specific time use codes for the activities mentioned earlier. It is possible that a few respondents may choose to report their active content creation in the category 'computer use for volunteering' (15-01-01). But that activity accounted for only 4 hours per person per year in 2016, far less than the 252 hours spent on user-generated content estimated earlier. Instead, it is likely that active content creation is tracked in leisure categories like 'socializing and communicating with others' (12-01-01) or 'computer use for leisure' (12-03-08).

According to the ATUS, household production time dropped slightly from 1,219 hours per person per year in 2006 to 1,208 hours per person per year in 2016. The steady time input is accentuated by technological stagnation. Household appliances like washing machines and vacuum cleaners were once new technologies that dramatically increased household productivity (Greenwood and Vandenbroucke 2005), (Greenwood et. al. 2005), (Ramey 2009). But these household appliances have been nearly universal in American households for decades. As a result, the real output of traditional household services has grown much slower than overall GDP. In other words, post-2000 slowdown is even worse when household production is included in output.

In general, it is difficult to value household production. It might be true that the market sector produces services that appear to be close substitutes for household production.[15] But a detailed examination shows that market output has different attributes than household production. Consider the case of cooking services. On the one hand, amateur cooks might be lower quality because they have less formal training than restaurant chefs. On the other hand, amateur cooks might be higher quality because they have more experience with their family's food preferences. As a result, the imputed value for household production depends enormously on the conceptual framework used and the empirical data used to impute hourly output. BEA's existing satellite accounts value household production time at the market wage for general purpose domestic workers (Bridgman 2016).

[15] A few household services are such close substitutes for market production that they are already included in the official GDP statistics. In particular, the currently published GDP statistics include an imputation for the "rent" earned by owner-occupied houses. In addition, the System of National Accounts 2008 also recommends imputations for home-produced goods (Section 6.32), do-it-yourself home improvement (Section 6.37), etc.

4.2 Valuing user-generated content in our experimental account

Some optimistic researchers might value user-generated content based on market wages. This methodology produces extremely high levels of estimated output. In 2016, American adults averaged 4.8 hours per week on content creation. If we valued that time at the market wage for programmers in the Occupational Employment Statistics, this works out to $2.5 trillion of labor input [(4.8 hours per week)*(52 weeks per year)*(250 million adults)*($40 per hour)]. In Table 1, we showed that the time devoted to user-generated content quintupled from 2006 to 2016. Hence, one might calculate that nominal output of user-generated content grew by $200 billion [($2.5-$2.5*0.2)/10] annually from 2006 to 2016. This growth is large enough to completely reverse the previously reported stagnation in household production.

However, we believe that the $2.5 trillion value is implausibly high. Cooking, cleaning and childcare are necessary for survival – so families must outsource them to the market sector if they do not produce them within the household. In addition, it is common for individuals to quit their market sector job or reduce their work hours in order to focus on cooking, cleaning and childcare. Therefore, it makes sense to use market wages to value those activities. In contrast, user-generated content is not a necessity and very few families outsource its production to the market sector. In addition, most user-generated content is created during relaxation time that would otherwise be devoted to television viewing, or other passive leisure.

Consumer surveys provide an upper bound on the value of user-generated content. The recent paper "Using Massive Online Choice Experiments to Measure Changes in Well-being" (Brynjolfsson, et. al. 2017) asks Internet users to self-report a value for various components of the Internet. Their current surveys do not cleanly distinguish between user-generated content and other digital content, but the total reported for social media is only $322 per year.[16] Across the entire United States population, that works out to $70 billion of consumer surplus. User-generated content is available without charge to potential viewers, so consumer surplus is an upper bound on the value for the user-generated content hosted on social media platforms. Of course, $70 billion in value is not small – but it is much much lower than the $2.5 trillion value estimated using market wages as a proxy.

We believe that the value of user-generated content should be based on hourly output for leisure activities which create incidental output. We do not have any general statistics on the incidental output from leisure activities. However, in previous work we calculated that television viewers "earn" approximately $0.69 cents of content for every hour they spent watching commercials (Nakamura, Samuels and Soloveichik 2017).[17] If we use the same $0.69 hourly output to value user-generated content, we calculate that Americans contributed $43 billion of labor inputs [(4.8 hours per week)*(52

[16] Taken from Table 7, column 3. For discussion purposes, we take the numbers as given, without adjusting for survey methodology. We note that social media is not exactly the same as user-generated content. Some user-generated content is hosted outside of social media, and only a portion of social media is user-generated content.

[17] In 2015, the average adult spent about 5 hours per day watching television and commercials accounted for approximately 30% of time – so aggregate commercial watching time was around 150 billion hours. In the same year, advertisers and marketers provided $108 billion of "free" television content.

weeks per year)*(250 million adults)*($0.69 per hour)]. In section 3.2, we calculated that another $13.2 billion of capital inputs. In total, we calculated that user-generated content provided $56 billion of value-added in 2016. Combing the labor and capital inputs over the time series yields average annual growth in the quantity index of value added for user-generated content of about 19% per year between 2006 and 2015.[18]

We believe that a $56 billion value-added for user-generated content is more plausible. In comparison, we calculated that households received $128 billion of "free" platforms and "free" professionally-generated content in our previous research (Nakamura, Samuels and Soloveichik 2017). That paper focused on measuring gross output of "free" digital content, and did not distinguish between labor inputs, capital inputs and intermediate expenses. But BEA's pre-existing I-O tables show that intermediate expenses represented approximately half of gross output for the internet publishing industry (NAICS 514). If the same ratio applies to amateur user-generated content, then the gross output of user-generated content was $112 billion in 2015. In other words, user-generated content accounts for approximately 47% of the value [112/(112+128)] from the Internet. Of course, $112 billion is not a trivial amount of money. But it is not nearly enough to reverse the recent stagnation in either official GDP or household production.

To be clear, our $0.69 hourly output number does not imply that Americans are wasting their time on user-generated content. Leisure activities directly create enormous utility for their participants. If demand for leisure is elastic, new innovations in leisure technology may result in rational individuals choosing to reduce their work hours (Bridgman 2016, Aguiar et. al. 2017). In addition, leisure time may also help people stay mentally healthy and thereby enhance their work productivity. In other words, commenting on Facebook can be an enjoyable activity that helps people wind down after a stressful day. This relaxation is valuable even if the comments generated are worth only $0.69 per hour to society. Conceptually, this is similar to a fisherman who catches only one fish – the day wasn't wasted even if the fish are cheap at the grocery store.

5. Conclusions and Next Steps

In this paper, we have provided preliminary and experimental measures of the production of user-generated content. Our basic premise is that it is important to include the production costs associated with user-generated content to fully account for the production of "free" content that is available via the internet. This early research is to present some exploratory statistics on the labor, capital and nominal value-added associated with digital user-generated content. In future research, we hope to construct a more detailed experimental account.

[18] We construct value added growth as a tornqvist index of the growth of the labor (hours) inputs and IT capital inputs, weighted by the cost shares. This assumes no TFP growth in the production of UGC.

Bibliography

Abraham, K., & Mackie, C. (2005). *Beyond the Market: Designing Nonmarket Accounts for the United States.* The National Academies Press.

Acemoglu, D., Makhdoumi, A., Malekian, A. & Ozdaglar A. (2017). Fast and Slow Learning from Reviews. NBER Working Paper 24046 http://www.nber.org/papers/w24046

Aguiar, M., Bils, M., Charles, K. & Hurst, E. (2017) Leisure Luxuries and the Labor Supply of Young Men.

Bridgman, B. (2016). Accounting for Household Production in the National Accounts: An Update, 1965–2014. *Survey of Current Business* 96 (February): 1-5

Bridgman, B. (2016). Engines of Leisure. https://www.bea.gov/papers/pdf/leisure1612.pdf

Bridgman, B., Dugan, A., Lal, M., Osborne, M. & Villones, S. (2012). Accounting for Household Production in the National Accounts, 1965–2010. *Survey of Current Business* 92 (May): 23–36.

Brynjolfsson, E., Eggers, F. & Gannamaneni, A. (2017) Using Massive Online Choice Experiments to Measure Changes in Well-being. Draft paper available at http://www.imf.org/en/News/Seminars/Conferences/2017/05/03/5th-statistical-forum

Greenwood, J. & Vandenbroucke, G. (2005). Hours Worked: Long-Run Trends. NBER Working Paper http://www.nber.org/papers/w11629.pdf

Greenwood, J., Seshadri, A., & Yorukoglu, M. (2005). Engines of Liberation. Review of Economic Studies 72, 109-133. http://www.jeremygreenwood.net/papers/engines.pdf

Jorgenson, D., Ho, M. & Samuels, J. (2017) Education, Participation, and the Revival of U.S. Economic Growth. NBER Working Paper 22453 http://www.nber.org/papers/w22453

Jorgenson, D., Nomura, K. & Samuels, J. (2016) A Half Century of Trans-Pacific Competition Price Level Indices and Productivity Gaps for Japanese and US Industries, 1955-2012, In *The World Economy: Growth or* Stagnation, 34-69. Cambridge, UK: Cambridge University Press

Landefeld, J.S., Fraumeni, B., & Vojtech, C. (2009). Accounting for Household Production: A Prototype Satellite Account Using the American Time Use Survey. *Review of Income and Wealth* 55, No. 2 (June): 205–225.

Nakamura, L., Samuels, J., & Soloveichik, R. (2017). Measuring the ”Free” Digital Economy within the GDP and Productivity Accounts. *BEA Working Paper*.

Ramey, V. (2009). Time Spent in Home Production in the Twentieth-Century United States: New Estimates from Old Data. The Journal of Economic History, Cambridge University Press, 69 (01) 1-47

von Hippel, Eric (2017). *Free Innovation*. Cambridge MA: MIT Press. Book draft available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2866571

Figure 1: Share of American Adults Online and Producing Content

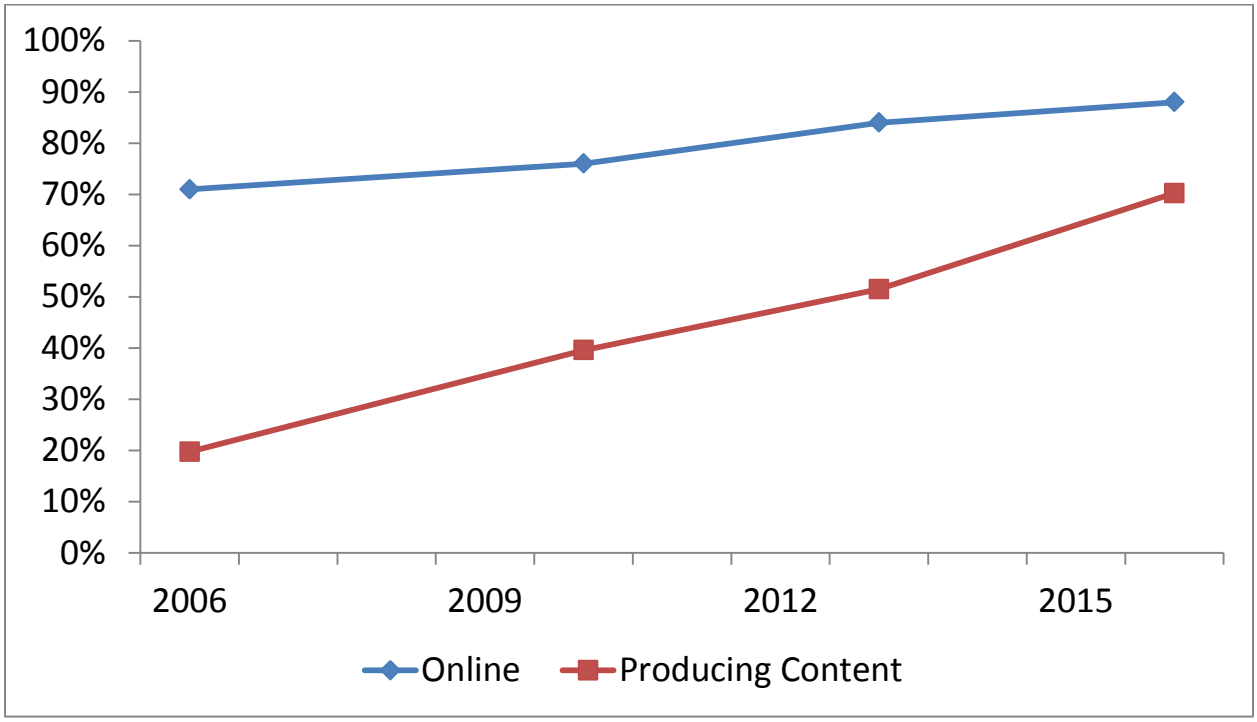

Figure 2: Share of Adults 18-34 Online and Producing Content

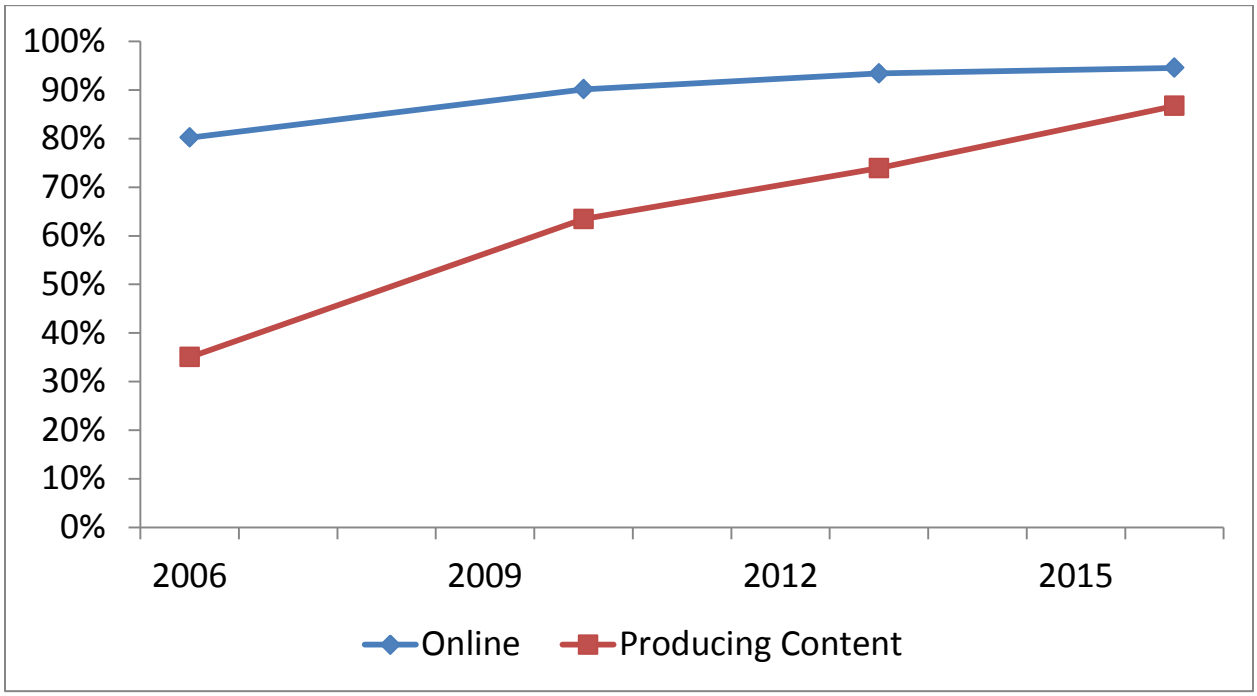

Figure 3: Share of Adults 35-54 Online and Producing Content

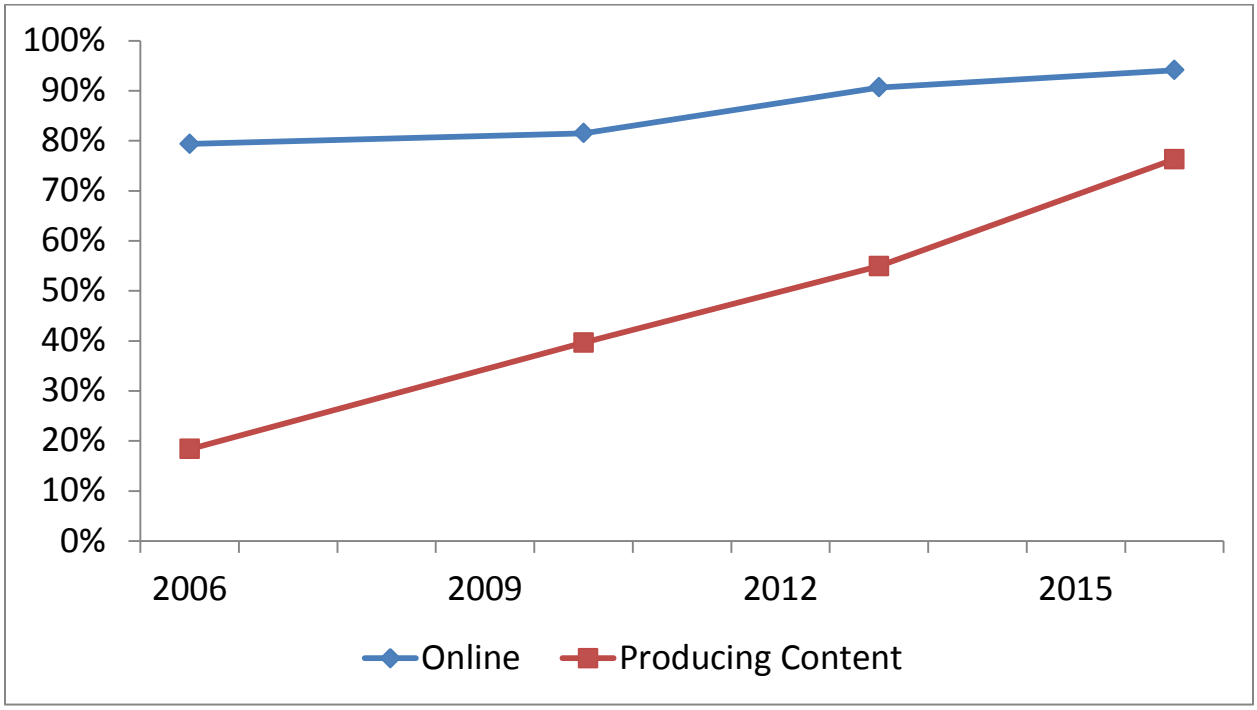

Figure 4: Share of Adults 55 or Older Online and Producing Content

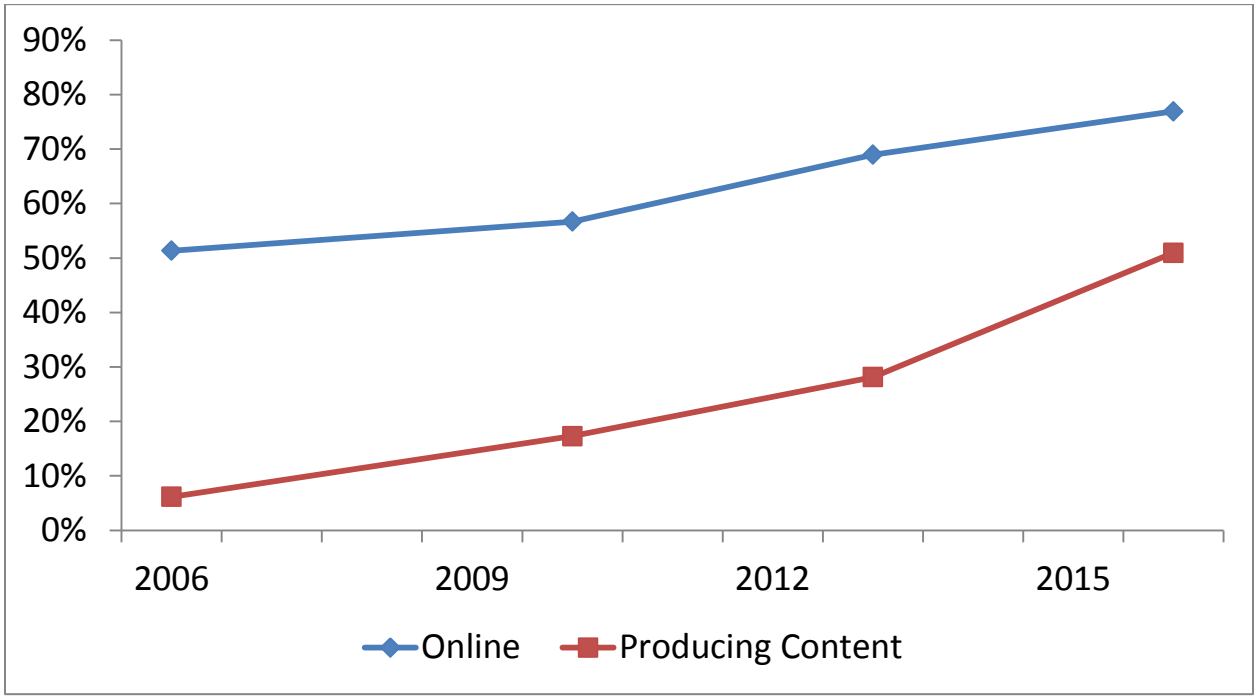

Figure 5: Male Share of Digital Population

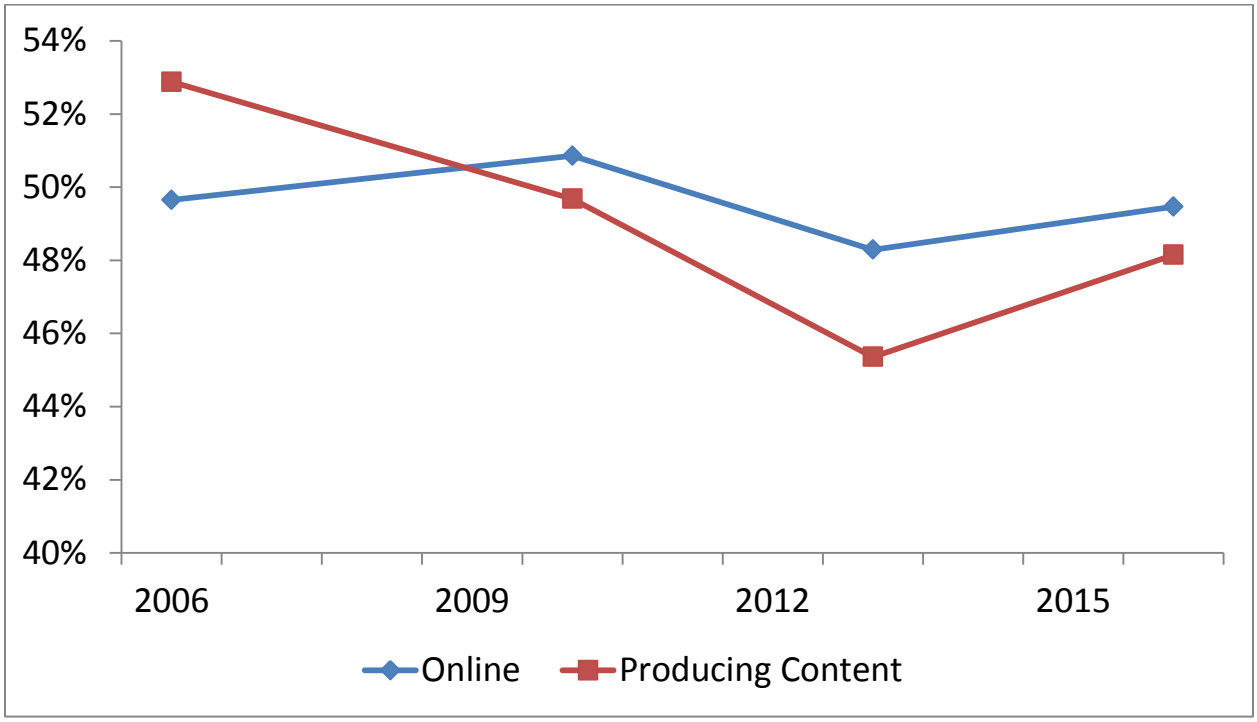

Figure 6: Millions of Adults Generating Content by Activity, 2010

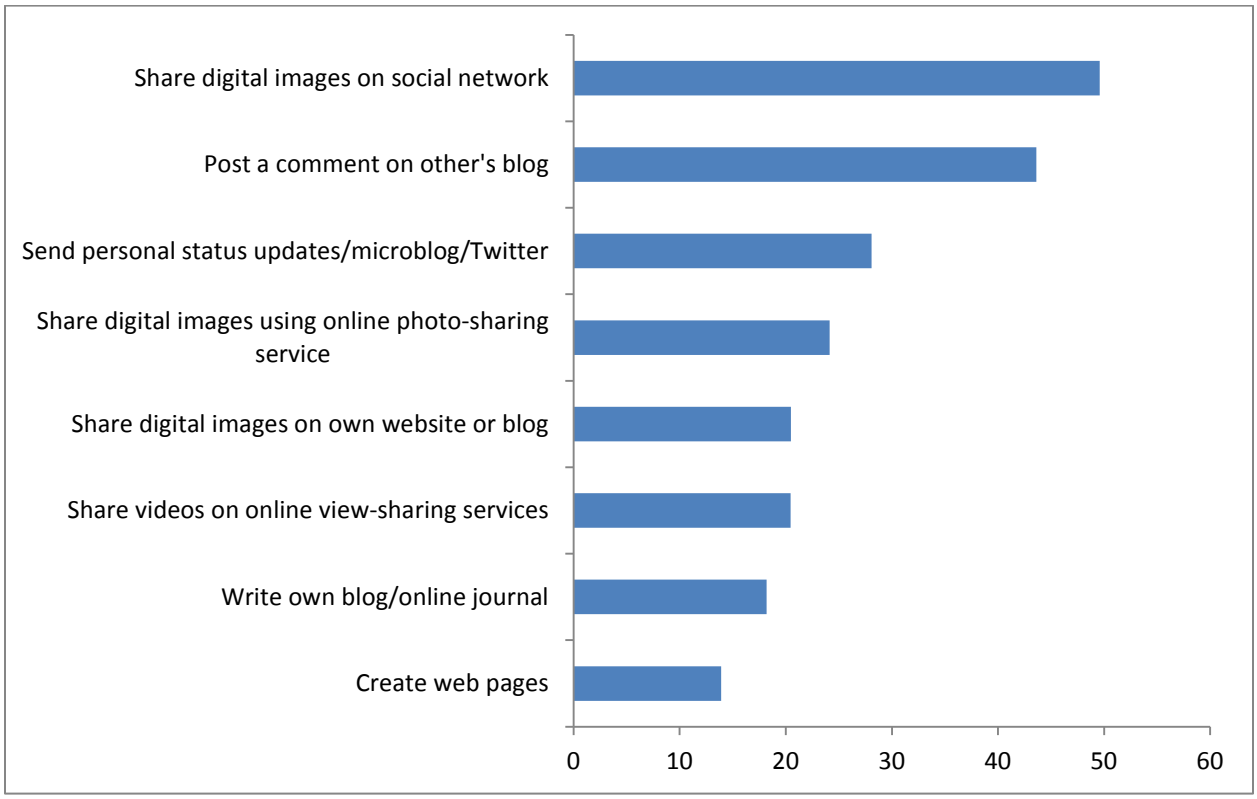

Figure 7: Millions of Adults Generating Content by Activity, 2016

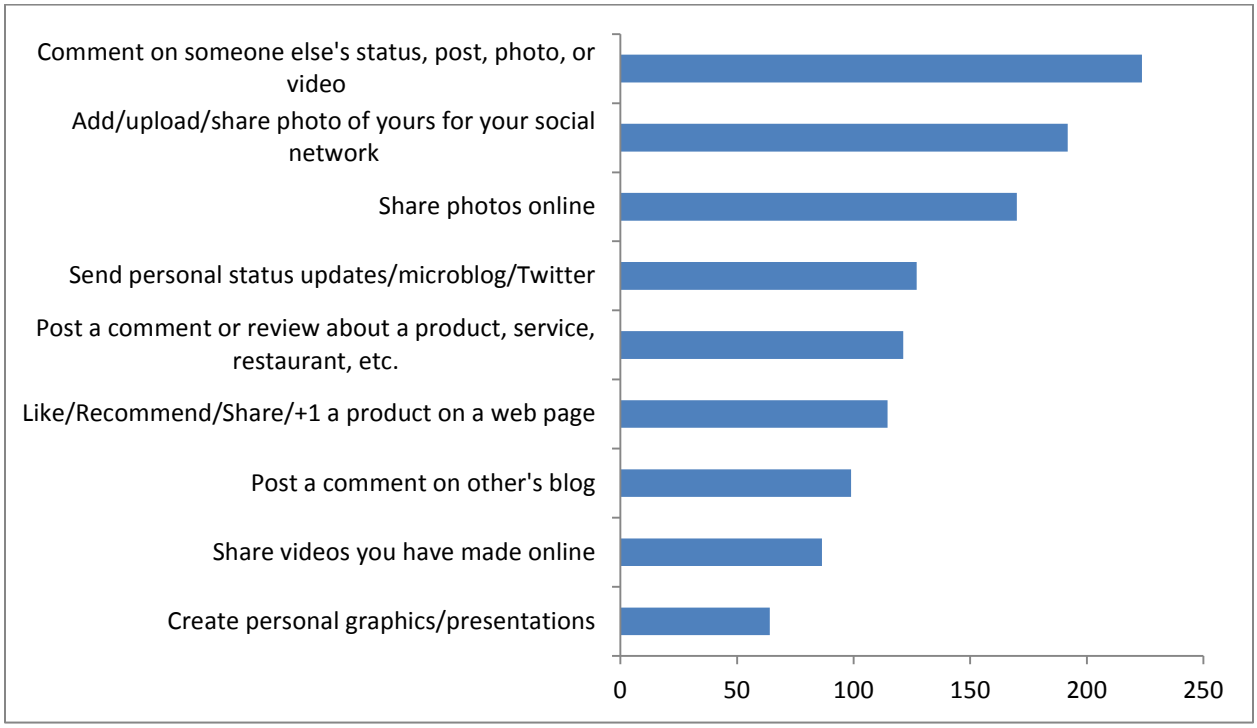

Table 1: Average Hours per Week Generating Content

		2006	2010	2016
Male	Age 18-24	2.1	5.2	5.3
Male	Age 25-34	1.4	4.2	6.3
Male	Age 35-44	0.7	3.4	5.4
Male	Age 45-54	0.6	2.0	3.5
Male	Age 55-64	0.4	1.5	3.2
Male	Age 65+	0.3	1.2	2.0
Female	Age 18-24	2.2	5.0	6.8
Female	Age 25-34	0.9	4.5	6.6
Female	Age 35-44	0.7	3.2	5.0
Female	Age 45-54	0.6	2.2	5.3
Female	Age 55-64	0.2	1.6	5.0
Female	Age 65+	0.1	0.9	2.9

Table A: Activity list 2016

Activities Associated with User-Generated Content are highlighted

Social network activities: Comment on someone else's status, post, photo, or video
Social network activities: Add/upload/share photo of yours for your social network
Social network activities: RSVP to an event for your social network
Social network activities: None of the above social networking activities
Shopping/purchasing activities: Compare products/prices while in store
Shopping/purchasing activities: Redeem a coupon/daily deal
Shopping/purchasing activities: Small purchases in person (gas, coffee, etc.)
Shopping/purchasing activities: None of the above shopping activities
Shopping/purchasing activities: Shop for products to buy online
Shopping/purchasing activities: Shop for products to buy in-person at a retail store
Shopping/purchasing activities: Download free apps/Software
Shopping/purchasing activities: Post a comment or review about a product, service, restaurant, etc.
Shopping/purchasing activities: Like/Recommend/Share/+
Shopping/purchasing activities: Purchase consumer services/products
Shopping/purchasing activities: Download paid apps/Software
Shopping/purchasing activities: Online banking
Communication activities: Send/receive personal email
Communication activities: Participate in a personal web-based group meeting
Communication activities: Participate in a work web-based group meeting
Communication activities: None of the above communication activities
Communication activities: Send/receive work email
Communication activities: Text messaging
Communication activities: Post a comment on other's blog
Communication activities: Send personal status updates/microblog/Twitter
Communication activities: Make/receive personal phone calls
Communication activities: Make/receive work phone calls
Communication activities: Make/receive personal video calls
Communication activities: Make/receive work video calls
Entertainment activities: Play a game
Entertainment activities: Listen to free streaming music (e.g. Pandora, Spotify Free)
Entertainment activities: None of the above entertainment activities
Entertainment activities: Watch videos/movies
Entertainment activities: Listen to streaming audio/Internet radio/podcast
Entertainment activities: Listen to music on this device, not on Internet
Entertainment activities: Watch television
Entertainment activities: Read a book
Entertainment activities: Download music
Entertainment activities: Hobbies
Entertainment activities: Listen to free or paid streaming music (e.g. Pandora, Spotify)
Graphics and image activities: Take Pictures

Graphics and image activities: Create personal graphics/presentations
Graphics and image activities: Print to a remote printer using an online service
Graphics and image activities: Create videos for work purposes
Graphics and image activities: None of the above Graphics & Image activities
Graphics and image activities: Store/scan photos
Graphics and image activities: Show or display photos in person
Graphics and image activities: Record Videos
Graphics and image activities: Share photos online
Graphics and image activities: Print wirelessly (WiFi) to a nearby printer
Graphics and image activities: Print to a remote printer using email
Graphics and image activities: Share videos you have made online
Graphics and image activities: Create work graphics/presentations
Information search activities: Search-other personal issues
Information search activities: Search-other work issues
Information search activities: Search on health
Information search activities: Read a magazine, newspaper, or periodical
Information search activities: Search on work finance/loans/investing/real estate
Information search activities: Search on personal finance/loans/investing/real estate
Information search activities: None of the above Information & Search activities
Cloud activities: Back up personal files/documents online
Cloud activities: Back up work files/documents online
Cloud activities: Cloud storage/sharing-personal files
Cloud activities: Cloud storage/sharing-work files
Cloud activities: Connect to a remote device/PC
Cloud activities: None of the above online cloud storage/sharing activities
Personal/productivity activities: Store/manage phone numbers/contacts
Personal/productivity activities: Manage tasks/to-do items
Personal/productivity activities: Download/use/update security/anti-virus software
Personal/productivity activities: Use an ad-blocking software app (e.g. AdBlock Plus)
Personal/productivity activities: None of the above personal/productivity activities
Personal/productivity activities: Collaborate on work files (e.g. Google Docs)
Personal/productivity activities: Collaborate on personal files (e.g. Google Docs)
Personal/productivity activities: Personal/household activities
Personal/productivity activities: Save/play voice memos (not voicemail)
Personal/productivity activities: Store/manage personal appointments/calendar
Personal/productivity activities: Store/manage work appointments/calendar
Personal/productivity activities: Finances/accounting
Personal/productivity activities: Write/manage text/notes/documents