Research Report

Economic Value of Data

FIRMS COLLECT A LARGE AMOUNT OF DATA BY ENGAGING HEAVILY IN THE COLLEC-
TION AND STORAGE OF ONLINE USER ACTIVITY VIA VARIOUS USER TRACKING TECH-
NOLOGIES. RECENT POLICY INITIATIVES AIM AT RESTRICTING THIS PRACTICE TO
PROTECT CONSUMER PRIVACY. WE STUDY EMPIRICALLY THE CONSEQUENCES OF SUCH
RESTRICTIONS FOR ONLINE PUBLISHERS, SUCH AS NEWS WEBSITES, BECAUSE THEY
STRONGLY RELY ON REVENUES THAT ARE GENERATED BASED ON USER DATA. WE
FIND A PRICE DECREASE OF CA. 30% FOR ONLINE ADS WHEN NO DATA FROM USER
TRACKING IS AVAILABLE. THE POTENTIAL REVENUE LOSS COULD BE MORE THAN
EUR 14 BILLION IN THE EU AND MORE THAN USD 27 BILLION IN THE US.

Rene Laub   Klaus Miller

Bernd Skiera

Motivation

According to the Internet Advertising Bureau
(IAB), in 2018, the online advertising industry
generated revenues of more than USD 107 bil-
lion in the US and more than EUR 55 billion in
the EU, with steady year-over-year growth of
more than 10%.

Transactions between publishers and advertis-
ers create the vast majority of these revenues.
Publishers, like the Financial Times or Spiegel
Online, make most of their money by selling ad
space on their websites to advertisers, who use
the ad space to display a specific ad to an indi-
vidual user of the publisher website. With
increasing data science capabilities, adver-
tisers draw on a large amount of data to per-
sonalize the content of the ad to the interests of
each individual user in real-time; thereby, they
increase the relevance of the ad for each user.
For that purpose, the online advertising indus-
try collects and stores a record of a user's activ-
ity on the Internet via various tracking technolo-
gies – with third-party cookies as the most pre-
dominant form. Third-party cookies are text
files that contain unstructured data about a
user's browsing history and are shared among
advertisers and publishers across websites.
Assume a user is browsing the Internet visiting
real estate websites, like immobilienscout24.de
in Germany. A small piece of text would be
saved in the third-party cookie. Advertisers can
access this data and deduct that this user might
be interested in a mortgage and target this user
with a matching ad.

With the growing discussion on the protection of
consumer privacy, the tracking of a user's
browsing history is under fire. Policy makers
have put forward regulation to restrict the collec-
tion, storage, and processing of user data as
introduced with the General Data Protection
Regulation (GDPR) in 2018 in the EU or the
California Consumer Privacy Act (CCPA) in 2020
in the US. Web browsers like Mozilla's Firefox,
Apple's Safari, and Google's Chrome have
already disabled or plan to disable tracking tech-
nologies, like third-party cookies, by default. This
development might have important economic
consequences for publishers, which rely heavily
on data-based online advertising income to
finance their free editorial content.

If online user tracking technologies are no
longer available, online advertisers (1) lose the
ability to profile users and personalize ad con-
tent, and (2) are no longer able to measure the
success of their online ads, e.g., by observing
click-through rates. As a result, advertisers'
willingness to pay (WTP) for displaying a specif-
ic ad to a specific user might decrease and, con-
sequently, also prices for ads and, thus, adver-
tising revenues of publishers might decrease.

While disabling user tracking technologies
potentially declines advertisers' WTP and, there-
by, leads to reduced ad prices, there are also
arguments that point in a different direction.

Disabling user tracking technologies might
increase ad prices by increasing competition
between advertisers. If user tracking is possi-
ble, data about a user's interests and prefer-
ences deduced from a user's browsing history
allows advertisers to segment the large amount
of online users and target only users that fit the
profile of an advertiser's target audience. Given
that not all advertisers are interested in the
same users, user data narrows down user seg-
ments and thereby decreases competition for a
specific user resulting in lower prices. Without
user data, the effect might reverse. More adver-
tisers could compete for the same users, lead-
ing to thicker markets, higher ad prices, and,
thereby, higher publisher revenues (Levin &
Milgrom, 2010).

So far, the potential effect of disabling user
tracking technologies is unclear, theoretical
predictions are mixed and only very little and
conflicting empirical evidence exists (Johnson
et al., 2020). We address this gap by empirically
investigating a unique data set of millions of
ad transactions to understand changes in ad
prices. We, thereby, inform policy makers and
industry participants about the monetary
consequences of the actions restricting user
tracking.

Description of Empirical Study

To assess the potential changes in publisher rev-
enues, we estimate the difference in prices of an
ad that are paid with and without user data. We,
therefore, examine the prices of more than 42
million ad impressions sold via a large European

ad exchange within a period of two weeks. These ad impressions account for approximately 5% of all ad impressions of the data provider during the observation period. Around 85 % of the ad impressions have associated user data available via a third-party cookie and around 15% of the ad impressions are without associated user data. The ads are shown to more than 1.3 million different online users and comprise 100 different online and mobile publishers, covering a broad variety of topics like cars, computer and technology, finance, games, health and lifestyle, or sports.

The average price, measured in cost per thousand (CPM), for an ad impression is EUR 0.63. With an average CPM of EUR 0.69, prices for ads with associated user data are higher than prices for ads without user data that have an average CPM of EUR 0.28. Thus, prices for ads with user information are EUR 0.41 (ca. 146%) higher than prices for ads without such information. Yet, this price difference cannot directly be interpreted as the increase that user data causes. The reason is that this difference does not account for other factors that impact differences between the prices, such as contextual data of the publisher, location of the user, content of the publisher website, or characteristics of the ad space (e.g., size or position).

We, therefore, consider these factors in a regression and also account for selectivity concerns. Selectivity concerns arise because the (un-)availability of user data is not random but a result of a user's deliberate choice of web browser or the installment of privacy management apps blocking user tracking. Therefore, users who do not allow for tracking might be systematically different to users who allow for tracking, e.g., in their preference for online advertising, and those (unobserved) differences might impact ad prices. If these unobserved differences influence the probability of having user information associated, then we will have an imbalance of these (also price influencing) factors between ads with and without user information. We, therefore, use augmented inverse probability weighting (AIPW). AIPW is a two-step procedure: We, first, estimate for each ad the propensity of having user information associated using a logistic regression. We, then, use in the second step the inverse of this propensity in a linear regression of ad prices on all observable ad price determinants. This weighting creates a balance in the unobserved difference between ads with and without user information.

Empirical Findings

Controlling for all observable ad impression price determinants, we estimate an average CPM price of EUR 0.64 for ad impressions with user data and EUR 0.44 for ad impressions without user data. As a result, user information yields prices that are EUR 0.20 (ca. 45%) higher. Stated differently, disabling user tracking could, therefore, lead to a reduction in ad impression prices of ca. 30%, as Figure 1 depicts. Given that the total share of ad impressions with user data accounts for around 85% of all ad impressions, the vast majority of ad transactions in the market could suffer from a severe price reduction. Assuming that the total advertising revenue numbers stated in the outset came from up to 85% of ad transactions with user data, the potential loss in the EU would be more than EUR 14 billion and in the US more than USD 27 billion.

We, further, investigate whether the potential price reductions differ between publishers. We observe the highest price reduction for publishers that provide content related to (1) cars and (2) computer and technology products. User data in these industries seems to be especially valuable. Publishers with content related to shopping and lifestyle products indicate the lowest price reduction. User data seems, therefore, to be especially valuable related to high-priced products (e.g., cars or computers). Advertisers could have a higher WTP for data of users interested in these products due to higher expe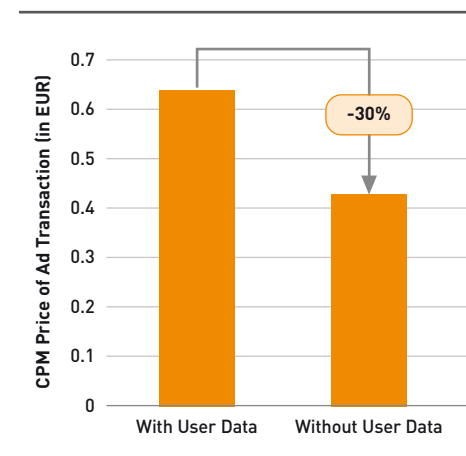cted profits compared to lower-priced products in the shopping and lifestyle area. User preferences for high-priced products are usually also more specific, compared to broader user preferences in lower-priced products, which could make user data more valuable in high-priced product environments.

Conclusion

Our study demonstrates the high economic value of user data. We find that prices for ad impression drop by ca. 30% when no user data is available. Since many companies base their business models on advertising revenues, our prediction could guide managers of these companies when designing new pricing schemes or switching to other business models. For example, companies could offer users the possibility of not collecting user information and in exchange charge a subscription fee that compensates for losses in advertising revenue. Our research results inform policy makers about the potential economic losses when balancing user privacy interest and interests of companies in the future.

References
Levin, J; Milgrom, P.:
Online Advertising: Heterogeneity and Conflation in Market Design.
In: American Economic Review, 100 (2010) 2, pp. 603-607.

Johnson, G. A.; Shriver, S. K.; Du, S.:
Consumer Privacy Choice in Online Advertising: Who Opts out and at What Cost to Industry?
In: Marketing Science, 39 (2020) 1, pp. 33-51.

Figure 1: Price Differences of Ad Impressions