

Making use of UGC in IO and Policy

Judy Chevalier

Yale and NBER

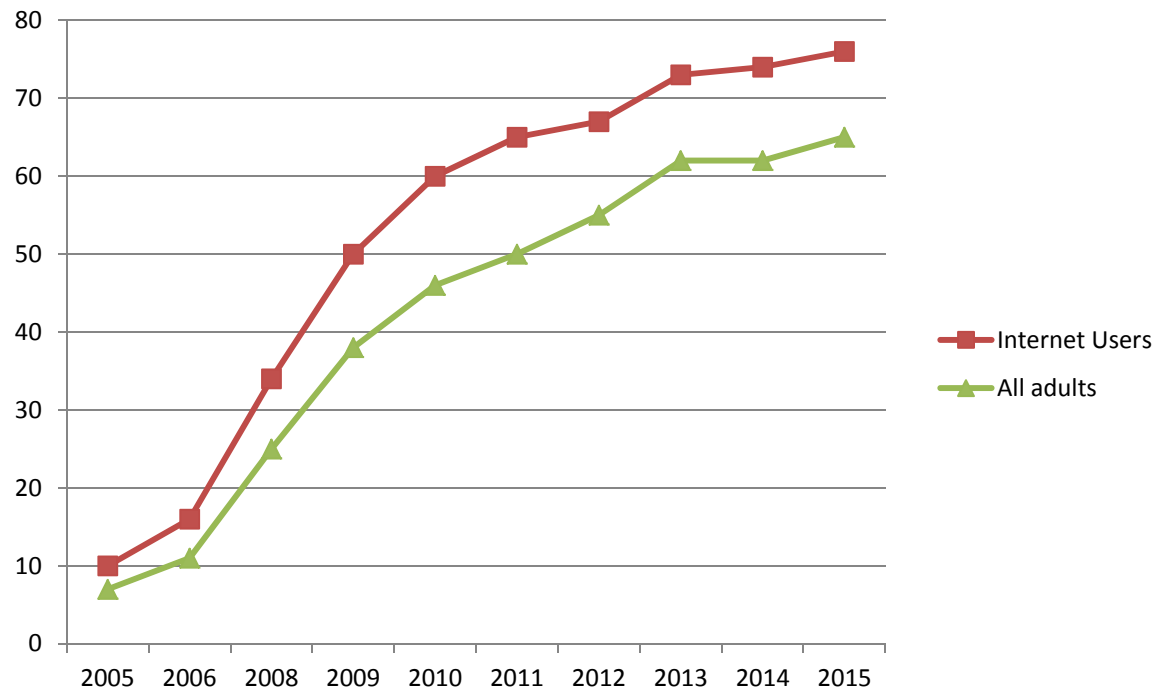
FTC Microeconomics Conference

Definition

UGC : any form of digital content that is produced by (typically) volunteer end users of an online service or website and (typically) publicly available for consumption by other users of the service or website.

Examples: videos, blogs, discussion forum posts, photos, product reviews.

UGC



- In 2004, Pew survey found 44% of Internet users had ever created online content of any sort.
- In 2013, found that 62% had posted or reposted photos online.

% of all American adults and Internet-using adults who use at least one social networking site (Source: Pew Research Center Surveys).

Existing literature in economics/marketing

- Large focus on product reviews in particular.
- Already a giant literature but a few areas of focus.
 - Do reviews causally impact sales?
 - Example Identification strategies: cross-platform comparisons (Chevalier and Mayzlin (2006)), IV (Shriver et. al. 2012), regression discontinuity((Luca (2011))), field experiments (Godes and Mayzlin (2009))
 - Who posts reviews and why?
 - See survey Berger (2015)
 - Are reviews systematically biased?
 - Reciprocity bias: Cabral and Hortacsu (2010), Fradkin et. al. (2015)
 - Promotional behavior bias: Mayzlin, Dover, and Chevalier (2014), Luca and Zervas(2015)
 - Positive selection: Hitt and Li (2008), Horton (2014), Nosko and Tadelis (2015)
 - Systematic review dynamics over time: Li and Hitt (2008), Godes and Silva (2011), Moe and Trusov(2011)
 - Design of review systems
 - Ursu (2015), Nosko and Tadelis (2015), , Dai et. al. (2014)

Limitations

- Large number of studies of review characteristics or testing whether reviews cause sales
 - Little on relationship between UGC and more traditional IO/Policy concerns
 - In Econ & Marketing (but not CS): Substantial focus on quantitative review summaries (stars)
 - Challenge: bridge unstructured textual data created by UGC and quantitative inputs to IO/policy analysis.

Possible research avenues

- How does existence/proliferation/nature of UGC *drive* variables of interest in IO/policy?
- How can UGC be used to *measure* variables of interest in IO/policy?

Possible research avenues

- How does existence/proliferation/nature of UGC *drive* variables of interest in IO/policy?
- How can UGC be used to *measure* variables of interest in IO/policy?

Market Structure

- Intuition from Bar-Isaac, Caruano, Cunat (2012):
 - More product information (lower search costs)
 - Superstar effect
 - Consumers more likely to find the “best product”
 - Encourages more-advantaged firms to choose broad designs
 - Niche effect
 - Consumers are more likely to buy better-suited products
 - Induce more firms to choose more niche designs
 - Effects of more information/reduced search costs on competition and markups ambiguous.

Empirical findings

- Relationship between existence of ecommerce platforms in general and product variety
 - Brynjolffson, Hu, and Smith (2010)
 - Quan and Williams (2015)
- Relationship between UGC specifically and product variety.
 - Luca (2011)

Impact of UGC on quality provision

- Quality improvements
 - Klein, Lambertz, Stahl (2015)
- Reviews and market outcomes in a multitasking framework.
 - Example: patient reviews of doctors

Competition among UGC platforms and competition between UGC platforms and other firms

- Does the existence of a UGC asset create competitive advantages in a transaction market?
 - The content is usually non-rival and non-excludable.
 - Freestanding UGC content providers entering the transaction role.
 - Yelp reservation system
 - TripAdvisor direct booking

Possible research avenues

- How does existence/proliferation/nature of UGC *drive* variables of interest in IO/policy?
- How can UGC be used to *measure* variables of interest in IO/policy?

Measurement of characteristics

- Inputs to demand analysis
 - We typically use introspection/data availability to choose characteristics to include in models.
 - UGC contains rich information about characteristics that consumers evaluate.
 - Or at least a selected set of consumers.
 - Closest paper: Archak, Ghose, Ipeirotis (2012)
 - Methodologically, though AGI examines impacts of reviews on sales, not characteristic measurement for demand estimation.

Measurement of characteristics

- Challenge: Bridge from textual content to quantitative data required by demand/IO models
- Possibility: Import tools from giant/growing CS literature on textual analysis.
 - Example: Latent Dirichlet Allocation
 - HB model- given a set of documents, learn the latent topics underlying the set and find words associated with topics.

Measurement of characteristics

- Challenge: Bridge from textual content to quantitative data required by demand/IO models
- Possibility: Import tools from giant/growing CS literature on textual analysis.
 - Example: Latent Dirichlet Allocation
 - HB model- given a set of documents, learn the latent topics underlying the set and find words associated with topics.

Second choice/Close competitor modelling

- Anything like second choice data helpful in demand system modeling (see BLP 1998)
- Textual analysis of UGC can be used for “brand associative network” (see Netzer et al 2012).
 - In reviews/blog posts about product x, what other products are referred to and how many times?

Measures of heterogeneity

- Harder: measurement of disagreement/heterogeneity of preferences over specific characteristics.
- Even harder: measurements of the distribution of the heterogeneity.

Possible research avenues

- How does existence/proliferation/nature of UGC *drive* variables of interest in IO/policy?
- How can UGC be used to *measure* variables of interest in IO/policy?