FESAC: Measurement of The Digital Economy

Patrick Bajari

VP and Chief Economist Amazon

12/15/2017

Outline

1. What is the Digital Economy
2. Productivity and Data Collection
3. e-Commerce Definitions
4. Quality Adjustments
5. Recommendations

New Technologies

- The Digital Economy is best understood as a set of technologies
- Examples: Internet, mobile phones, apps, gps/mapping, machine learning, cloud computing, virtual reality, advanced logistics, autonomous driving, artificial intelligence/robotics
- These technologies will be deployed in almost all industries

  -Agriculture: Tractors controlled via GPS/Satellites to more efficiently plant seed and spread fertilizer. Optimized with ML.

  -Automotive: Tesla uses connectivity, data, ML/AI, advances in computing for autonomous driving

  -Retail: Target and Shipt. Personal shoppers fulfill online orders made on your cell phone from local retailer in hours.

Economic Importance

1. Connect and communicate with customers
    - -Improves information that customers have about product
    - -Allows firms to continuously improve and personalize the product
    - -Creation of liquid marketplaces
2. Efficiency and the scientific method
    - -ML, AI, etc… are fundamentally about building models using data
    - -Remove magic numbers, rules of thumb and other heuristics from software
    - -Replaced with rational decision making and the scientific method

Diffusion

- Much like the tech wave of the 1990's, new technologies take time to diffuse
- Most important barrier- management
- Need to move from a world where core decisions are guided by science instead of heuristics
- How can you manage software guided by machine learning and OR without fluency in these fields?
- Business schools are seeing MBA programs shrink at the margin and analytics grow
- This tech wave may diffuse more slowly than the last

Misconceptions

- The digital economy is not an industry
- These technologies will be adopted by all industries
- Much like the diffusion of faster CPUs, cheaper storage and better software in the 1990s
- “Tech” firms are early adopters because they tend to have more scientifically literature management
- The benefits of improved connectivity and efficiency are common to all industries

Misconceptions

- The economic benefits when we look back in 20 years may be relatively small in tech

  -Deadhead loads in trucking

  -Data driven HR to improve retention and productivity

  -Food wasted between field and consumer in agriculture

  -Enabling small businesses to reach consumers at scale (400K US jobs associated with merchants on Amazon alone)

  -More rational decisions in millions of firms using the scientific method to improve worker productivity

Role for agencies

- What technologies are being used and by whom
- Improve measures of the capital stock and use of technologies at the firm level
- Are firms hiring workers with training in ML, Stats, OR
- Relate these technologies to productivity at the firm level
- We are at risk of making bad policy
- Agencies should document the diffusion of these technologies and their relationship to output and efficiency

E-commerce definitions

- Treating e-commerce separate from the rest of retail feels very 1990's and is at odds with how industry is evolving
- All of retail is quickly innovating to use these new technologies in ways that blur a simple division between online and offline

  -Using voice search in my auto's nav system to find store, gas or restaurant

  -Ordering coffee in my Starbucks app to avoid the long lines

  -Using virtual reality to visualize and size a piece of clothing to purchase from a retailer

  -Pop up store fronts: Indochino (made to measure suits) or Warby Parker (glasses). Showroom with inventory managed using an e-commerce model.

  -Instore pickup and returns for online sales (e.g. Home Depot reported this year that more than 40% of the company's U.S. online sales are picked up in store)

  - Using delivery service to get goods from physical stores within hours. Amazon Restaurant, Instacart, Shipt, Google Express.

  - Mobile Apps for faster in store checkout (Target) and returns (Walmart)
- Traditional retailers have seen the largest and fastest growing online sales, boosted by their physical locations: according to eMarketer Walmart, the Home Depot, Best Buy, Macy's, and Costco are among top 10 online retailers and have all seen double-digit annual percentage growth in online sales (Walmart was up 47% , Best Buy 30%, and Home Depot 28%).

E-Commerce Definitions

- With Amazon, we are unable to replicate Census's e-commerce sales figures using internal or publicly available sources
- Michael Mandel has shown problems with employment estimates
- Diversion ratios between "e-commerce" and "offline" retail are close to -1 in Forrester Data
- Billion price project shows identical retail "online" and "offline" prices
- Comscore, NPD, Argus, and numerous surveys- clear that customers multihome
- Fails common sense tests for a separate industry

Quality Adjustments

- Adapt technique from the private sector and modernize data collection infrastructure
- Data is increasingly ubiquitous

  -Augment the CEX with an app that allows agencies to capture credit card, Amazon purchase history, Google search history data

  -Go behind firms firewalls to build quality adjusted price indices or demand models

  -Scrape the web for product data.

  -Create regressors from text using Natural Language Processing

  -Use image processing for pictures

Quality Adjustments- Hedonics

- Imagine you have daily data on prices
- Let *i* denote a product and *t* time
- $p_{i,t}$ an unbalanced panel of prices
- $x_{i,t}$ time varying product features

Hedonic Price Regressions with Quality Adjustments

- Model:

$$\ln(p_{i,t}) = \alpha + x'_{i,t}\beta_t + \gamma'_i f_t + \varepsilon_{i,t}$$

- Regressors $x_{i,t}$ with time varying implicit prices β_t
- Vector of unobserved product attributes γ_i.
- We can interpret f_t as time varying implicit prices on unobserved attributes
- $\varepsilon_{i,t}$ idiosyncratic error
- Ideally with daily data

Hedonic Price Adjustments

- Berry (1994), Berry, Levinsohn and Pakes (1995), Nevo (2001)
- Unoberved product attributes are often quantitatively as important as observed attributes
- We should also use more modern data sources NLP and Image Processing
- Factor Augmented Regression- Bernanke et al(2004), Bai(2009) and Hastie et. al.
- Benkard and Bajari (2005) apply this model to computers to adjust price indexes
- We can use the same specifications for demand systems e.g. quantity or discrete choices are the dependent variable

Hedonic Price Adjustments

- Pooling data across retailers could help to adjust for the unobserved quality associated with new technologies deployed by different retailers (outlet substitution bias)
- Micro-econometric models can learn what is important to customers rather than an approach based on applying ad hoc judgments and dated distinctions such as “online” and “offline”
- We can learn these using observed characteristics or by relating latent factors SKU/retailer differences
- Amazon would volunteer to use a decade of daily price data on millions of products to demo this for the agencies

Summary and Recommendations

1. Think of the Digital Economy as a technology, not an industry
2. Define what technologies are likely to have economic importance
3. Focus on measuring how these technologies are diffused through the economy through software, employment patters capital stock, etc…
4. Modernize your data collection
    -Behind company firewalls
    -App for collection of consumer panels
6. Quality Adjustments and Hedonics
    -Microeconometric and panel models
    -Unobserved Product Attributes
    -Data from NLP, Image Processing and new data sources
7. Amazon is willing to help