How Big Data has Changed Agriculture

Keith Coble
@DrKeithHCoble
coble@agecon.msstate.edu
Where I am Coming From?

• Career researching agricultural risk
  – Example weather weighting crop insurance loss experience
• Spent time on Capitol Hill as a Chief Economist
• Council on Food, Agricultural and Resource Economics ‘Big Ag. Data Steering Committee’
• Agricultural & Applied Economics Association conference on “Big Ag Data”
• Board member for the Ag Data Coalition – farm groups, agribusiness firms, and Universities engaged to manage data
Adoption of precision technology is coming on rapidly

The Ag Funder Agtech Investing Report identifies approximately $1.4 billion total investment in two categories during 2015.

Source: David Schimmelpfennig, Farm Profits and Adoption of Precision Agriculture, ERS Report 217
Advancing U.S. Agricultural Competitiveness with Big Data and Agricultural Economic Market Information, Analysis, and Research

• Big data is associated with precision agriculture, but is much larger than on-farm data alone:
  – On-Farm Precision Ag
    • Fertilizer, planting, crop protection, yield monitors
    • Government sources of weather, satellite, soil maps
  – Supply Chain
    • Food safety traceability, sustainability verification, supply chain efficiency
  – Consumer Demand
    • Scanner data, social media

Keith Coble, Terry Griffin, Mary Ahearn, Shannon Ferrell, Jonathan McFadden, Steve Sonka, and John Fulton
Adoption of precision technology is coming on rapidly

The Future of On Farm Data

Estimated Amount Of Data Generated By The Average Farm Per Day

Source: OnFarm, BI Intelligence Estimates, 2015
Agtech Financing 2010 – 2015

- 2010: $0.4
- 2011: $0.5
- 2012: $0.5
- 2013: $0.9
- 2014: $2.4
- 2015: $4.6
Making Farm Management Sexy Again
One Vision of the Future

• Monsanto executives are seeking to reposition the company as a business built on data science and services, as well as its traditional chemicals, seeds and genetic traits operations, Chief Technology Officer Robert T. Fraley told Reuters in an interview.

• “We transformed from industrial chemical company to a biotech company, then to a seeds company,” Fraley said. “Now, we’re transforming again.”
Big Data is Defined by four Vs

- **Volume** the new paradigm is that the data are too big to move so the analytics must move to the data.
- **Velocity** enables managers to respond to events as they are happening, rather than after the fact.
- **Variety**, - sensors on equipment, controllers, RFID tags, drone imagery, etc. Often not structured or scientifically complete
- **Big Data Analytics** are the tools and techniques to turn big ag data into useful information.
Turning Big Ag Data into Knowledge

• Consider a world where we know:
  – Variety planted, soil type, planting population, fertilization, weather, yields, other inputs on millions of acres
  – Non-randomized design, some inaccurate data
  – Not from an experimental design

• Can we learn this messy data – absolutely!
  – Billion are being invested in learning from this data
  – Analysis will be far more complex than experimental plots
  – Yield = f(variety, soil, planting population, fertilization, weather, other inputs)
A Vision of Agricultural Research in the Future

(1) Research and Experiment Data
Shared and organized in a useable fashion

(2) Program Administrative Data

(3) Private Precision Data Re: Weather, Inputs, and Outputs

(4) Surveys that Collect Missing Data
Privacy protected such that research is achieved while avoiding private data release

Data held by farms & agribusiness firms that is a new avenue for research using machine learning etc.

Government statistical agencies maintaining important benchmark surveys & captures important information like farm financials and prices.
Challenges to Our Profession

1. **Maintaining, storing, and aggregating big data**—While the value proposition of big data appears immense, such data is messy, requires standards, and is valuable in disaggregate and aggregate form.

2. **Tracking food safety**—It is essential to track, monitor, and eradicate threats to our food, feed, fiber, and fuel supply. To build a safer and more economically viable U.S. food supply, we need to harness the possibilities of big data.

3. **Matching products with markets**—Consumer decisions reflect a complex matrix of priorities. Increasingly, consumers are considering health, environment, affordability, and quality as they purchase food & related products. It is difficult to track materials through complex supply-chain processes.

4. **Ameliorating survey gaps**—New approaches may be possible, using precision data that reduce the need for some survey-based data. This information could be stored behind a secure firewall, shared, and updated cost-effectively.
5. **Mitigating asymmetric market information**—Policy issues regarding data ownership, market power, and privacy are not well understood.

6. **Advancing evidence-based policy**—Evidence-based policy is policy informed by rigorously established objective evidence. It can be advanced with the onset of more seamless and robust survey and big data information that can also be disaggregated in secure environments.

7. **Ensuring researcher access to big data**—Big data may be a significant asset for supporting productivity gains and policy improvements. However, researchers need access to the data. Data markets may be one way of enabling access; strong partnerships may be another.
What will the Ag Data Scientist of 2020 look like?

• Trained to use geo-spatial data
• Hard science need to become comfortable with analysis of non-experimental data
• Can working in multi-disciplinary teams
• Ag Data Scientists will be trained in machine learning algorithms, data cleaning, & use of less structured data
WHO BENEFITS FROM THE INFORMATION?
Farmer Fears About their Data

1. The people I share my data with are making money from my data & I am not.
2. The people I share my data with will know every thing about me and use it against me.
3. My data may fall into the hands of the government or an environmental group and they will use it against me.
The Ag Data Transparency Evaluator

- Website created by a non-profit corporation backed by a consortium of farm industry groups, commodity organizations and ag technology providers in order to bring transparency, simplicity, and trust into the contracts that govern precision agricultural technologies.
- Based upon the foundation laid by the Privacy and Security Principles for Farm Data (the Data Principles), the Ag Data Transparency Evaluator is a process by which ag technology providers voluntarily submit their ag data contracts to a simple, ten question evaluation.
- Only companies receiving approval are allowed to use the “Ag Data Transparent” seal.
- [http://www.fb.org/agdatatransparent/](http://www.fb.org/agdatatransparent/)
What is Big Ag Data Worth?

To You
• Depends on whether
  – Does it allow you to make better decisions
  – Faster decisions
  – Manage more
  – You are prepared to use it
  – The quality of the data

To Others
• To know more about you
• To compare you to peers
• To aggregate to large scale
• To merge with other data to study
So who owns this image?
Who OWNS “Ag Data”?

• How do you OWN property?
  – Right to POSSESS
  – Right to USE
  – Right to ENJOY
  – Right to EXCLUDE OTHERS FROM
  – Right to TRANSFER
  – Right to CONSUME or DESTROY

Source: Paul Goeringer – University of Maryland
Market Power Issues

• Many of the agribusiness firms involved
  – Machinery companies
  – Seed/genetics
  – Fertility
• Different business models
  – Tie the farmer to your company and data apps
  – Not make data a profit center
• Questions being asked
  – If input firm has significant knowledge of my farm, can they price discriminate?
  – Can markets arise that allow farmers to sell access to data?
Bias and root mean squared error for sampling methods across various sample sizes

<table>
<thead>
<tr>
<th>Random Average</th>
<th>Samples are drawn randomly from all farms in the data. Weights are set to 1/Z.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random All</td>
<td>Samples are drawn randomly from all farms in the data. Acreage weights.</td>
</tr>
<tr>
<td>Random I-state</td>
<td>Samples are drawn randomly from Iowa, Illinois, and Indiana. Acreage weights.</td>
</tr>
<tr>
<td>Random C-belt</td>
<td>Samples are drawn randomly from corn-belt states: Iowa, Illinois, Indiana, Kentucky, Michigan, Minnesota, Missouri, Nebraska, and Ohio. Acreage weights.</td>
</tr>
<tr>
<td>Random S-rank</td>
<td>Samples are drawn randomly within each state. The number of farms from each state is determined by the percentage of acreage in that state. Acreage weights.</td>
</tr>
</tbody>
</table>

Source: The Potential Implications of ‘Big Ag Data’ for USDA Forecasts by Jesse Tack, Keith H. Coble, Robert Johansson, Ardian Harri, and Barry J. Barnett
SOME CASE STUDIES OF WHERE WE ARE HEADED
U.S. Crop Insurance

- A public/private delivery system
- Assessing price and yield risk accurately is extremely difficult
- Risk quantification is inherently data intensive
- Significant rate improvements over time, but within county variation poorly understood
Crop Insurance Rating

- Corn insurance (Revenue Protection)
- 697 counties in Corn Belt States
- 11 years balanced panel: 2005–2015

- Soil productivity: NCCPI
  - National Commodity Crop Productivity Index (NCCPI), USDA/NRCS
  - Rating of soil production capacity of dry-land crops: 0 to 1
  - soil properties, landscape features and climatic characteristics
  - 10-meter → County-level
Data

11-year Average Loss Cost Ratio

11-year Average Base Rate

NCCPI
Results

Estimated $f(\text{Soil})$

OLS Spatial HAC Model

- Blue curve: Step function
- Red curve: Linear splines
- Comparative static $f(\text{Soil})$
- Relative loss cost
- Reject $H_0$:
  Soil is irrelevant
Will Big Ag Data allow Greater Profitability and Sustainability?

- Sub-field management
- Verification of practice or sustainability
A Few Really Important Points from Economics

• What is the objective when choosing
  – Which crop to plant?
  – How much nitrogen to apply?
  – Which variety to plant?

• A common myth that is often repeated
  – Maximizing yield = Maximizing profit

• Reality
  – Farmers tradeoff maximizing profit subject to avoiding risk and other constraints
Policy Analysis Going Forward

• Much of the big data advancement has been in predictive models – non-causal models
• Policy analysis needs causal models
  – Some new big data analytics will be adapted to causal models
• The role of Government statistical data collection
  – Equalizing asymmetric information
Some Thought on the Future of Big Ag Data

• Infrastructure and technological solutions will be found that address the limits to rural broadband access.
  – Economists can provide estimates of the value of addressing this infrastructure.

• Progress will be made in developing more sophisticated nutrient, irrigation and environmental management, farm management
  – Validation
  – mechanisms to certify sustainability practices
Some Thought on the Future of Big Ag Data

• Policy, market, and contractual issues regarding data ownership and the privacy standards associated therewith will need to be developed.

• Greater use will be made of spatial and high-volume data analytical techniques in agricultural research.
  – This will require changes in curriculum and a demand for applied economists trained in these techniques.
Some Thought on the Future of Big Ag Data

• The ability to conduct research in this arena will require access to volumes of data controlled by others.
  – Researchers will need to have proven value to the gatekeepers of the data and be able to maintain the confidentiality of spatial data.
  – Because of volume and security concerns researchers will need to come to secure data enclaves that allow access to analysis but restricts access to the data.
Thank You