Farmers' willingness to participate in a big data sharing program: A study of Saskatchewan grain farmers

A Thesis Draft

Submitted to the College of Graduate and Postdoctoral Studies

In Partial Fulfillment of the Requirements

For the Degree of Master of Science

In the

Department of Agricultural and Resource Economics

University of Saskatchewan

by

Madeline G. Turland

PERMISSION TO USE

Head of the Department of Agricultural and Resource Economics
University of Saskatchewan
3D34 Agriculture Building, 51 Campus Drive
Saskatoon, Saskatchewan S7N 5A8
Canada

Dean
College of Graduate and Postdoctoral Studies
University of Saskatchewan
116 Thorvaldson Building, 110 Science Place
Saskatoon, Saskatchewan S7N 5C9
Canada

ABSTRACT

Turland, Madeline G. M.Sc. University of Saskatchewan, Saskatoon, July 2018. *Farmers' willingness to participate in a big data sharing program: A study of Saskatchewan grain farmers*

Supervisor: Dr. Peter Slade

Big data in crop agriculture is information collected by sophisticated machinery at the farm level, as well as externally generated data, such as field satellite imagery. Although some of this data is useful to individual farmers, much of it has little value to the farmer that collects it. Capturing the true value of big data comes when it is aggregated over many farms, allowing researchers to find underlying bio-physical and economical relationships.

We conduct a hypothetical choice experiment to analyze farmers' willingness to share data by asking farmers in Saskatchewan whether they would participate in a big data sharing program. The choice tasks varied the type of organization that operated the big data program and included financial and non-financial incentives.

Heteroscedastic and random effects probit models are presented using the data from the survey. The results are consistent across models and find that farmers are most willing to share their data with university researchers, followed by crop input suppliers or grower associations, and financial institutions or equipment manufacturers. Farmers are least willing to share their data with government. Farmers are more willing to share data in the presence of a financial incentive or non-financial incentive such as comparative benchmark statistics or prescription maps generated from the data submitted.

Checks for robustness and heterogeneity indicate there is no self-selection bias into the survey, and no heterogeneity in the results for financial incentive and farm revenue. A latent class logit model determines the farmer population may be heterogenous in their willingness to participate in a big data sharing program, but homogenous in their ordering of preferences for organization, financial incentive, and non-financial incentive. In addition, demographic variables are not related to class membership.

ACKNOWLEDGMENTS

I would like to express my thanks and appreciation to my supervisor, Dr. Peter Slade. This thesis is a testament to the countless hours I spent in his office, and the many more he spent reading my (often crude) writing. Peter encouraged me to attend conferences and present my work when I had the opportunity. His guidance, support, and mentorship through this process was invaluable to my success. I hope many more students will have the privilege to study under him. I would also like to thank my committee members Dr. Richard Gray, Dr. Eric Micheels, and Dr. Stuart Smyth for their valuable contributions and timely feedback.

A special thanks goes to my parents, Robert Turland and Eve Geisler, for their support through this process, and for the help and guidance they have given me through my life. Their life lessons and encouragement made this process much easier. I would also like to thank my partner, Devin Serfas, who believed in me through this entire process. His emotional support gave me the determination to keep working, and the time he spent reading and editing improved the quality of my work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Modern precision agricultural technologies have the ability to generate detailed farm and field level datasets, mapping inputs and outputs to the sub-field level, for large crop operations. This data is not only useful to farmers but has the potential to transform technological innovation and production in the agriculture industry. This type of data is a subset of what is commonly referred to as big data. Historically, big data is defined as datasets so large they cannot be processed by traditional software (Cox and Ellsworth, 1997). However, that definition has evolved and the term "big data" now has many interpretations. Commonly, big data is described by three Vs; volume, velocity, and variety. In this thesis, I focus on data generated by precision agriculture equipment on crop operations in Saskatchewan. Across farms, this type of data rarely differs in velocity or variety, but by aggregating it over many farms, the volume can be greatly increased.

Big data is used in almost every industry, and researchers are finding ever more innovative ways of extracting value from it. However, in crop agriculture, much of the value of big data is being left on the table (or in the field). Regional analysis is rarely performed because farm level big data is seldom aggregated (Poppe, Wolfert, Verdouw, and Renwick, 2015). A dataset aggregated over many farms becomes valuable for researchers attempting to uncover underlying bio-physical and economical relationships that could help generate improvements in yield, environmental stewardship, automation, other areas related to agricultural production, and where research dollars should be spent in the future.

Ownership of big data in agriculture is a murky question. Some argue that the companies that make the equipment that captures the data (such as John Deere) have some legal claim to the data, while others say that producers retain full ownership (Sykuta, 2016). When farmers purchase equipment, they receive an implied license for the life of the vehicle to operate the vehicle, but the copyright for the software that runs the equipment remains with the manufacturer (Lyseng, 2018; Schemper, 2014; Wiens, 2015). It is unclear where the generated data lies on the spectrum from farmer ownership to manufacturer ownership. One way to aggregate farm level data is to set up a big data sharing program, and have farmers voluntarily participate. Achieving high participation rates is a primary goal when constructing a big data program because the value of the big data sharing program increases with the number of farmers participating as a result of network externalities. However, if farmers place a high value on privacy, they may be reluctant to share information about themselves and getting them to voluntarily participate could be difficult. This thesis estimates farmers' willingness to share big data, who they are willing to share it with, and what can incentivize them to do so.

The purpose of this thesis is to determine the key factors that influence grain farmers' willingness to participate in a big data program in Saskatchewan. I conducted a survey of grain farmers in Saskatchewan, asking a series of hypothetical choice questions. Survey respondents were asked if they would participate in a big data program under varying conditions. The choice questions varied; (a) the organization running the program, (b) the financial incentive for participation, and (c) the non-financial incentive for participation. The specifics of the survey are detailed in the data section of this thesis. A heteroscedastic probit model and a random effects probit model are estimated. A latent class model is also presented to check for heterogeneity in the results.

The results are consistent across models. The institution responsible for coordinating the big data program has a particularly important effect on participation rates. Farmers are most willing to join a program run by university researchers, followed by crop input suppliers or grower associations, then equipment manufacturers or financial institutions. Farmers are least willing to participate in a program run by government. Positive financial incentives increase farmer participation rates, while monetary contributions for participation (negative financial incentives) have no impact on participation rates. The financial incentives studied are small compared to farm revenue, supporting the existence of a privacy paradox. Non-financial incentives, such as benchmark statistics or prescription maps generated from data submitted, increases farmers' willingness to share data, but less so than financial incentives.

There has been no previous work analyzing farmers' willingness to share data, although some work has been done measuring privacy behaviours in the general population. Privacy preferences do not appear to change after the individual has received education about privacy policies and potential risks (Olsen, Grundin, and Horvits, 2005; Debatin, Lovejoy, Horn, and Hughes, 2009). If this holds true for farmers, then they will not be induced to share data through social coercion methods such as advertising campaigns. Other methods, such as financial or non-financial incentives for participation, must be used.

Often, people will state strong privacy preferences, but appear to violate these preferences for remarkably small rewards. This is known as the privacy paradox (Athey, Catalini, and Tucker, 2017; Norberg, Horne, and Horne, 2007; Barnes, 2006). This paradox is clearly demonstrated by

Athey, Catalini, and Tucker (2017). In the paper, students at the Massachusetts Institute of Technology (MIT) were induced to violate their stated privacy preferences for a slice of pizza. Relative to their actions, people tend to overstate their privacy preferences when directly asked. The privacy paradox could inform the results of this thesis; the incentives required to induce farmers to participate in a big data program may be surprisingly small.

This thesis adds to the body of work surrounding privacy and big data by analyzing farmers in Saskatchewan. This research provides a foundation for the construction of a big data sharing program for agriculture in the future. Understanding what organization would be most successful in establishing a big data program, and the effect of different incentives on participation rates will make the construction of the program more likely, and less costly. Future gains in productivity and efficiency in agriculture might be data driven, as most of the low hanging productivity gains have already been realized. Canadian agriculture must be willing to innovate to keep pace with a growing world population. A big data program allows researchers to do just that.

The first section below is further background surrounding issues related to big data and data management in agriculture. Section three outlines the survey, and data used for the statistical analysis. Section four presents the methods used. Section five shows the results, and section six presents a discussion and conclusion. Appendices 1 and 2 present full results tables, and a more detailed discussion of the method.

BACKGROUND

Big data is an abstract concept with many fluid definitions. The first use of the term big data was by Cox and Ellsworth (1997), who defined a dataset as "big" if it was too large to be processed by traditional software. Since that time the definition of big data has evolved to include many types of data that are "big" in different ways. George, Haas, and Pentland (2014) suggest that the fine-grained nature of the data defines big data; it does not matter how many individuals are in the dataset, but rather how much you know about each individual. Cukier and Mayer-Schoenbergen (2013) argue that big data is about learning things from a large body of information that was invisible in a smaller set. Trujillo, Kim, Jones, Garcia, and Murray (2015) describe the three V's of big data: volume, velocity, and variety. These characteristics reveal the challenges of collecting and working with big data; large amounts of everchanging data, requiring real time collection and analysis.

Until recently, the cost of data storage was so high that collecting and using big data was prohibitively costly in most situations. With falling costs of storage and analysis, big data is becoming a more widely used tool. Private companies are seeing the benefits of investing in data and competing to be the first in their industries to innovate. As the cost of data storage continues to fall, big data will become an ever more accessible tool for industry and government (Trujillo et al, 2015).

Working with big data brings challenges for researchers. The sheer number of observations in the dataset could make identifying important relationships difficult (George, Haas, and Pentland, 2014; Fan, Han, and Liu, 2014). Working with many observations means standard errors for any analysis will be low, and almost any relationship could be found to be statistically significant, when in fact no causal relationship exists. Researchers are faced with the prospect of sorting through large numbers of statistically significant relationships to determine the economically significant ones. In addition, there is a danger of overfitting models because of the sheer number of available variables (Fan, Han, & Liu, 2014). Data quality may be lower in big datasets as it becomes increasingly difficult to clean and curate data as the datasets increase in size. This should not be a significant problem however if there is at least some accuracy in the data. Some inaccuracies can be tolerated in exchange for the benefits that come with such large datasets (Cukier and Schoenbergen, 2013). Big data holds the potential to uncover population patterns and heterogeneities that are unexpected, and subtle relationships that cannot be found with conventional data.

In crop agriculture, big data describes datasets created by sophisticated machinery and software that quantify inputs and outputs at a micro level. Other types of data not generated by farm equipment such as satellite imagery are also included in big data. Big data is part of the precision agriculture revolution that aims to increase automation and productivity. Precision agriculture is the application of technologies to manage spatial and temporal variability associated with all aspects of agriculture production (Pierce and Nowak, 1999). Big data informs precision agriculture by providing the farmer with detailed information about their land and input use, allowing farmers to make better decisions and use their precision agriculture equipment more efficiently. An example of big data use in agriculture is mapping profits on a field in terms of input use and yield and converting areas of the field that consistently show negative profits to conservation (Coble, 2018). Other examples include benchmarking, sensor deployment and analytics, and predictive modelling (Wolfert, Ge, Verdouw, and Bogaardt, 2017). Big data can also be used to measure environmental degradation. Big data applications in farming are not strictly about primary production but play a major role in improving the efficiency of the entire supply chain and alleviating food security concerns (Wolfert, Ge, Verdouw, and Bogaardt, 2017).

Whatever the promised benefits, advances using precision agriculture and big data may be insignificant when it comes to increased production, as weather continues to be the most important factor. The benefits of big data and precision agriculture might be oversold to farmers, as many precision agriculture techniques have yet to be proven, and it is uncertain whether their promised benefits will ever be realized (Bronson and Knezevic, 2016). However, farmers must percieve at least some benefits from the use of precision agriculture techniques as the use of this technology is rising. In the experimental choice survey conducted for this study, 75% of respondents use yield monitors, 94% use GPS guidance, 77% use soil sampling, 29% use variable rate technology, and 56% use automatic section control.

In 1996, Steven Sonka and Karen Coaldrake considered what precision agriculture and advanced communication technologies would mean for farms. They imagined the possibility of a "Cyberfarm" (Sonka & Coaldrake, 1996). These operations would be capable of capturing and analyzing farm level data. They imagined a network of farmers sharing information using advanced communication technology. Although not all elements of the Cyberfarm have been realized, significant advancements have been made in communications, information sharing, and marketing. However, there remain challenges facing big data in agriculture. Currently, farm data is rarely shared, analyzed by intelligent software, or combined for regional analysis (Poppe, Wolfert, Verdouw, and Renwick, 2015). For the industry to be revolutionized, seamless data integration systems that farmers believe in must be put in place. The eventual impact of big data within the agricultural sector likely will require both organizational and technological innovation (Sonka, 2014).

It is unclear where the benefits of big data in agriculture will flow, however Sonka (2016) determines that consumers will be the ultimate beneficiaries as technological advancements tend to lower prices for consumers. In contrast, Bronsen and Knezevic (2016) predict that most of the benefits from big data will flow to a small number of large agriculture companies. This is a direct result of the oligopolistic nature of the agriculture industry. A small number of large firms control input supply to a large number of small farmers. Farmers could also see some benefits from big data use depending on how the big data market is ultimately structured. The reality is that many of the benefits that come from using big data and the distribution of those benefits remain unknown. Researchers do not yet know what they will be able to accomplish using big data in agriculture. The advances can not come if the data is unavailable, so the construction of a big data sharing program is the first step to big data research in agriculture.

There has been little work studying farmers perceptions of big data. Boyer, Engleking, and Gudas (2015) find that farmers have a positive view of big data, yet also value traditional farm management tools over more advanced technologies. Their study found that few participants indicated high awareness of data security and other risks, and increased concern about data security was not associated with age or education. Farmers' perceptions of big data are informed by the marketing tactics employed by companies that sell big data services.

Big data technologies use information collected by precision agriculture technologies. It follows that the adoption path of big data technologies will be similar to that of precision agriculture technologies. The adoption of precision agriculture technologies has been slower than predicted, but some producer characteristics have been linked to higher rates of adoption. Farm size is positively correlated with adoption, while age is negatively correlated (Tamirat, Pedersen, and Lind, 2018; Deberkow and McBride, 2003; Larson et al., 2008). In addition, computational literacy and education are positively correlated with the decision to adopt (Deberkow and

McBride, 2003; Larson et al., 2008). Looking at the non-agriculture private sector, factors affecting a firm's willingness to adopt big data technologies include the firm's human resources, technology resources, and management support (Sun, Cegeilski, Jia, and Hall, 2018).

Adoption of technologies not only depends on the individual considering adoption, but on the characteristics of the technology itself. Technologies that require the farmer to acquire additional knowledge to operate (such as variable rate technology) have lower adoption rates than those that can be integrated using existing knowledge (such as GPS guidance) (Miller, Griffin, Bergtold, Ciampitti, and Sharda, 2017). Firms in all sectors are more willing to adopt if the perceived benefits from the technology are high, and the cost of adoption is low (Sun, Cegeilski, Jia, and Hall, 2018). As the availability and variety of hardware and software needed to collect and analyze big data increases, farmers will be more willing to adopt, and there will be increased public sector initiatives and business ventures in the agricultural sector (Kamilaris, Kartakoullis, and Prenafeta-Boldú, 2017).

Farmer participation in a big data sharing program is akin to adopting a new technology and depends on the value the farmer will extract from participation. As a big data sharing program becomes larger, it not only becomes more valuable, but it also becomes more widely known among farmers. Products that gain value from a large userbase benefit from network externalities (Katz and Shapiro, 1985). Being first to the market could bring advantages in terms of establishing a loyal following among farmers, and an environment that lacks competition for market share. However, those that choose to enter the market later can observe and learn from the first-to-market's model and develop a better product. Product compatibility also becomes relevant when considering competition between database companies. The ease with which consumers can switch between companies will undoubtedly affect the market structure of data collection in agriculture (Katz and Shapiro, 1986). Perhaps there is a role for government to ensure compatibility between systems, allowing for competition in the marketplace.

The organization that ultimately invests in the construction of a big data program will determine the structure of the database within the constraints (if any) the government has provided, including who have access to the data, and how the data will be used. Bronson and Knezevic (2016) say "the use of large information sets and the digital tools for collecting, aggregating, and analyzing them … has the potential to wade in on long-standing relationships between players in food and agriculture."

There has been no previous work analyzing farmers' willingness to share data, although some work has been done measuring privacy behaviours in the general population. Privacy preferences are found to be fixed for individuals and do not change when individuals are educated about data sharing risks or the absence of risks (Debatin, Lovejoy, Horn, and Hughes, 2009; Olsen, Grundin, and Horvitz, 2005). These studies apply to personal data of people in the general population. Farmers' privacy preferences may be different than the general population, and farm data is business data rather than personal data. However, most farms are family run, and the business data is confounded with personal data making a clear separation between the two difficult. It is unknown whether farmers' treat their farm data as personal or business, and if that has any impact on data sharing preferences.

Relative to their actions, people tend to overstate their privacy preferences when directly asked. This is known as a privacy paradox (Norberg, Horne, and Horne, 2007; Barnes, 2006). An example of the privacy paradox is found in Athey, Catalini, and Tucker (2017). Students surveyed at the Massachusetts Institute of Technology (MIT) were found to relinquish data quite readily when incentivized to do so, even though their stated privacy preference may be strong. The privacy paradox highlights the importance of experiments in privacy research, as surveys may not adequately capture privacy preferences. Evidence of this privacy paradox is found in the results of this thesis, however, this thesis is based on data collected through a survey, and so the results are still subject to a hypothetical bias.

Pavolotsky (2013) identifies a unique problem with sharing big data. The value of big data may lie in identifying secondary uses of the data that are unimagined at the time of collection. When consent to share data is obtained, it applies only to those uses of the data that are conceivable. Keeping and using data for unimagined purposes stretches the limits of meaningful consent. This is illustrated in the Facebook Cambridge Analytica scandal that broke in early 2018. The Facebook profile data of up to 87 million users was collected and sold to Cambridge Analytica (Bloomberg, 2018). Cambridge Analytica then used the data to target voters with hyper-specific appeals, potentially having an impact on election results. Before the scandal broke, Facebook users would have been unaware that their data may have been used to influence election results. Perhaps they would have chosen to update their privacy settings to keep their information private had they been aware.

In agriculture, there are many stakeholders that see the value in constructing a big data program that encompasses data from many farms. Government has vested interests in greater agricultural production and higher farm incomes. University researchers want to expand the type of research they are able to undertake and conduct current research more accurately. Equipment manufacturers and crop input suppliers are continually attempting to innovate new products and improve services for farmers. Financial institutions want detailed information on a farmer's operation, so they can better evaluate their ability to repay loans. Grower organizations are looking for better ways to serve their industries. Each of these organizations could conceivably decide to pioneer a big data sharing program as all would see benefits. This research attempts to determine which organization would generate the highest participation rates among farmers, and what incentives are most effective in inducing them to do so.

DATA

The data come from a hypothetical choice experiment. The survey was administered online by Kynetec, a market survey company, to grain farmers in Saskatchewan from October 10 to November 20, 2017. Respondents were offered $10 in compensation for completing the survey. Due to lower than normal response rates, compensation was raised to $20 on November 1, and to $30 on November 8. Out of 561 respondents, 344 were compensated with $10, 129 were compensated with $20, and 88 were compensated with $30. Payments are controlled for in the analysis, and do not have a statistically significant impact on the results.

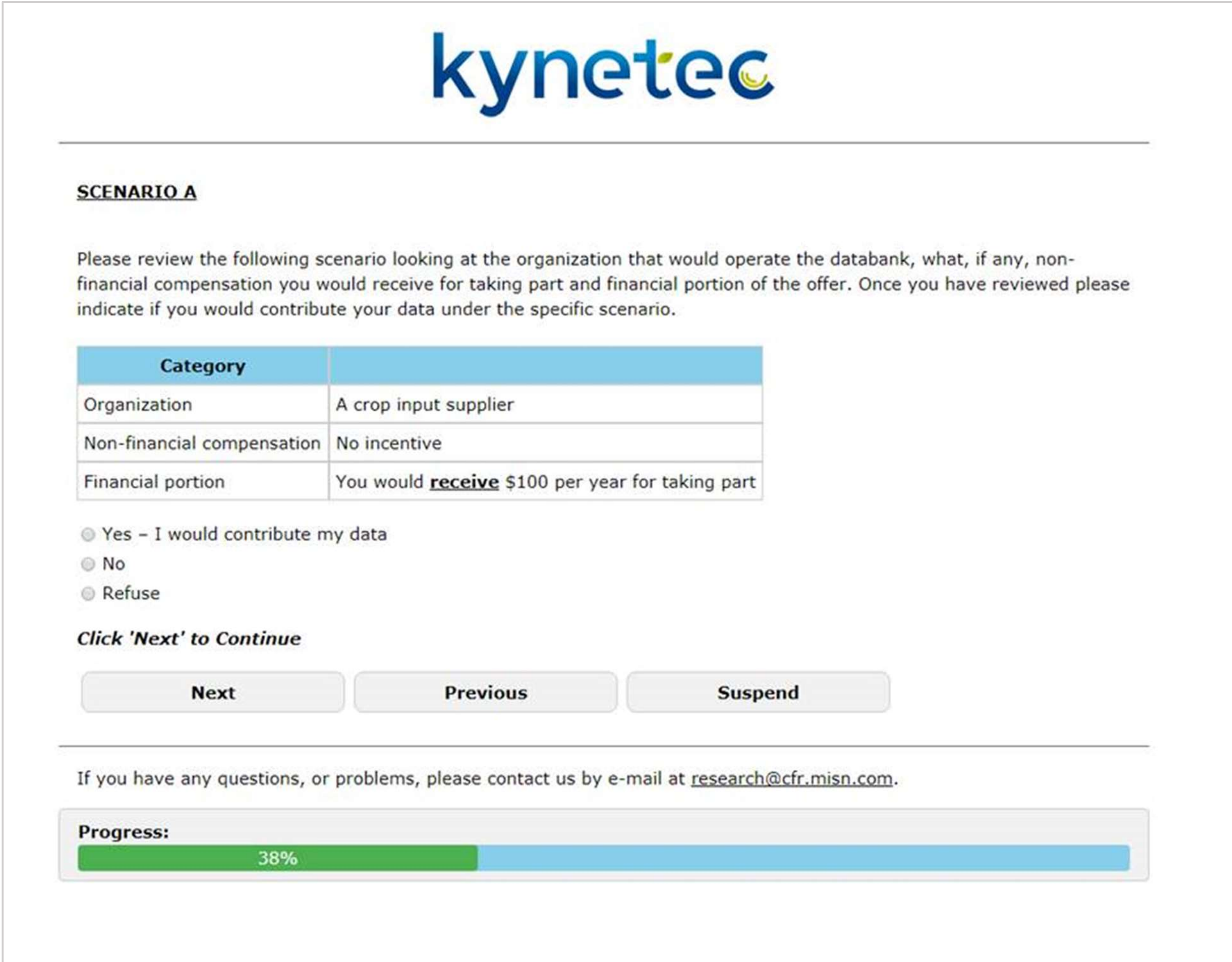
The survey asked farmers about their use of precision agriculture technology, attitudes towards privacy, technology use and farm management, and sociodemographic information. In addition, respondents were asked a series of choice questions. The choice questions asked if respondents would be willing to participate in a big data program under specific conditions.

The choice questions presented a set of scenarios that varied; (a) the organization running the big data program, (b) the financial incentive for participation, and (c) the non-financial incentive for participation. Table 1 shows the organizations and incentives that were included. One option from each category was chosen to formulate each choice question, and each respondent was asked to evaluate twelve separate choice questions. A screenshot showing an example choice question is presented in figure 1. Preceding the choice question set, the respondents were provided with an information script briefly explaining the rationale behind a big data program. The script also asks respondents to assume no transaction costs for program participation. A screenshot of this is shown in figure 2.

Table 1: Organizations and incentives studied

Organization	Financial Incentive	Non-Financial Incentive
1. University Researchers	1. -$50	1. None
2. Crop Input Suppliers	2. $0	2. Prescription maps based on the data submitted. Depending on the data submitted these could be for fertilizer, seed, fungicide, or other inputs.
3. Grower Associations	3. $50	
4. Equipment Manufacturers	4. $100	
5. Financial Institutions		3. Yield and input use benchmarks. For example, “of the farms in your area, your yields are in the 50th percentile while your fertilizer use is in the 75th.”
6. Government		

Figure 1: Example question from the choice question set



kynetec

SCENARIO A

Please review the following scenario looking at the organization that would operate the databank, what, if any, non-financial compensation you would receive for taking part and financial portion of the offer. Once you have reviewed please indicate if you would contribute your data under the specific scenario.

Category	
Organization	A crop input supplier
Non-financial compensation	No incentive
Financial portion	You would **receive** $100 per year for taking part

- Yes – I would contribute my data
- No
- Refuse

Click 'Next' to Continue

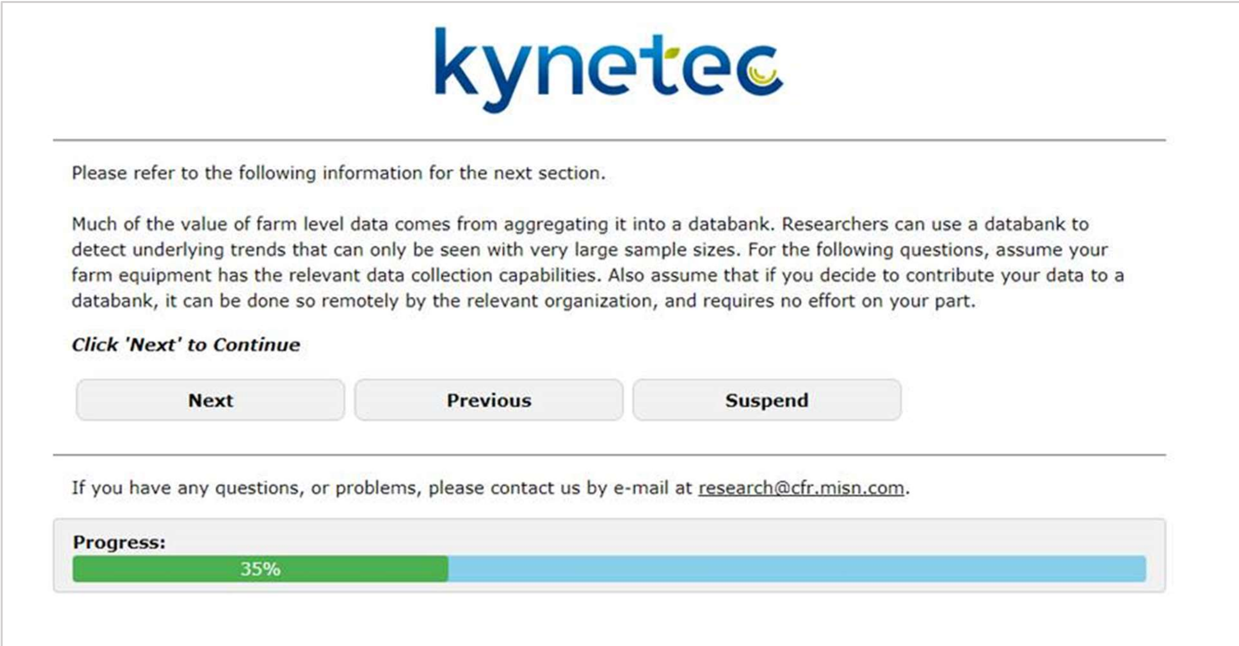
Next **Previous** **Suspend**

If you have any questions, or problems, please contact us by e-mail at research@cfr.misn.com.

Progress:
38%

Screenshot from the online survey conducted by Kynetec.

Figure 2: Information script preceding choice question set



Screenshot from the online survey conducted by Kynetec.

The survey design was pseudo-random. With the options studied, there are seventy-two unique combinations of organization, financial incentive, and non-financial incentive. The seventy-two unique scenarios were divided into six groups of twelve, ensuring sufficient variation in organization, financial incentive, and non-financial incentive within each group. Six versions of the survey were created, each asking one of these groups of twelve questions. Each respondent randomly received a version of the survey to answer. An approximately equal number of responses was received for each version of the survey.

In addition, respondents were asked to evaluate eleven statements by their level of agreement on a scale of one (low level of agreement) to five (high level of agreement). These statements attempt to capture farmers' attitudes towards privacy, technology use, and farm management. The first three statements measure privacy attitudes, the following four statements measure technology use attitudes, and the final four statements measure farm management attitudes. Since multiple statements are attempting to capture the same thing, we would expect the results within each subset of questions to be highly correlated. The statements, their mean responses, and standard deviations are shown in table 2. These questions are referred to as the attitude question set in this thesis.

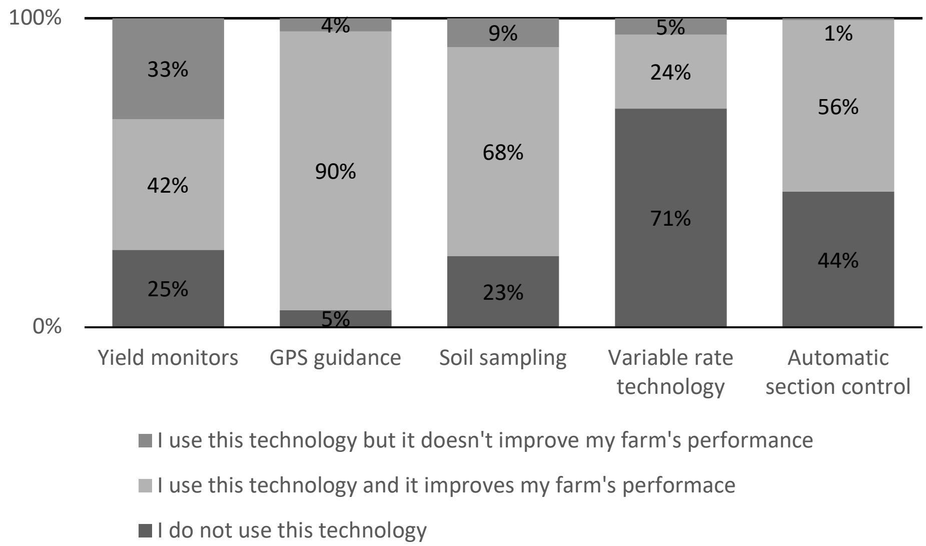
Table 2: Responses to attitude questions set (respondents rate their level of agreement on a scale from one to five)

	Mean	St. Dev.
Privacy		
Privacy is important to me	4.0	1.00
I would be put at a disadvantage if other could access info about my farm	3.1	1.10
I feel comfortable sharing information about my farm	3.2	0.95
Technology use		
I like to have the latest technology	3.4	1.00
I find new technologies easy to use	3.3	0.99
New technology is more hassle than it is worth	2.5	1.02
I am getting maximum use out of available tech on my farm	3.3	1.04
Farm management		
I have implemented new techniques that have been recommended	3.6	0.87
I am proactive in seeking advice	3.9	0.85
Precision ag will transform agriculture over the next 20 years	4.0	0.97
I know better than others how to manage risk on my farm	3.6	0.89

To gauge farmers' current use of modern technology, respondents were asked about their use of yield monitors, Global Positioning System (GPS) guidance, soil sampling, variable rate technology, and automatic section control. Respondents answered, "I do not use this technology", "I use this technology and it doesn't improve my farm's performance", or "I use this technology and it improves my farm's performance". The responses to these questions are detailed in figure 3. GPS guidance is the most widely adopted technology, while the fewest number of farmers have adopted variable rate technology. Yield monitors are highly adopted, however 44% of farmers who have adopted them say their farm performance hasn't improved as a result. These questions are referred to as the technology use question set in this thesis.

The five technologies studied in this thesis are all precision agriculture technologies, however not all of them utilize big data in a meaningful way. GPS guidance and automatic section control are mechanical technologies that don't require outside analysis for decision making. However, they are complementary to big data technologies as they must be adopted for big data technologies to be effective. Yield monitors and variable rate technology generate big data, and require outside analysis to be done for value to be extracted from their use. The survey asks questions about all of these technologies as the responses could indicate the speed at which a farmer adopts new technology.

Figure 3: Responses to technology use question set



The average age of those surveyed was 56.1 (standard deviation of 11.0). This is similar to the average age of farm operators in Saskatchewan in 2016, which was 55 years (CANSIM table 004-0017). Women are underrepresented in the survey; they make up just 5.5% of respondents while 24.9% of farm operators in Saskatchewan were female in 2016 (CANSIM table 004-0017). Of those surveyed, 5% did not have a high school diploma, while 23% had a university degree. In contrast, only 7.2% of farmers in Saskatchewan had a university degree in 2011 (CANSIM table 004-0110). In the survey, 58% of farms were incorporated, 51% used the services of an agronomist, and 26% used the services of a financial analyst.

Annual sales revenue was collected as a categorical variable. A comparison of survey data and data collected from Statistics Canada can be found in table 3. Of those surveyed, 3% reported annual sales revenues less than \$100,000, 29% reported revenue between \$100,000 and \$499,999, 27% reported revenues between \$500,000 and \$1 million, 20% reported revenues between \$1 million and \$2 million, 4% reported revenues between \$2 million and \$3 million, and 5% reported revenues greater than \$3 million. A significant portion (12%) of respondents refused to disclose their annual sales revenue. This distribution differs from that found by Statistics Canada (CANSIM table 004-0006). This discrepancy might be partially explained by differences in the farm population surveyed. Statistics Canada includes all farmers in their measure, including livestock, specialty crop, and hobby farms. These farms tend to be smaller and generate less revenue, skewing the distribution downward. In contrast, the survey conducted for this study included only grain farmers in Saskatchewan. In addition, the pool of farmers that Kynetec samples from is skewed towards larger operations.

Table 3: Proportion of Saskatchewan farms by revenue range

Revenue Range	Survey (grain farmers in SK)	Statistics Canada* (All farmers in SK)
<$100,000	3%	43%
$100,000-$499,999	29%	35%
$500,000-$999,999	27%	12%
$1 million - $2 million	20%	7%
$2 million - $3 million	4%	3%**
>$3 million	5%	

** CANSIM table 004-0006*
***The final two categories are combined in StatCan data*

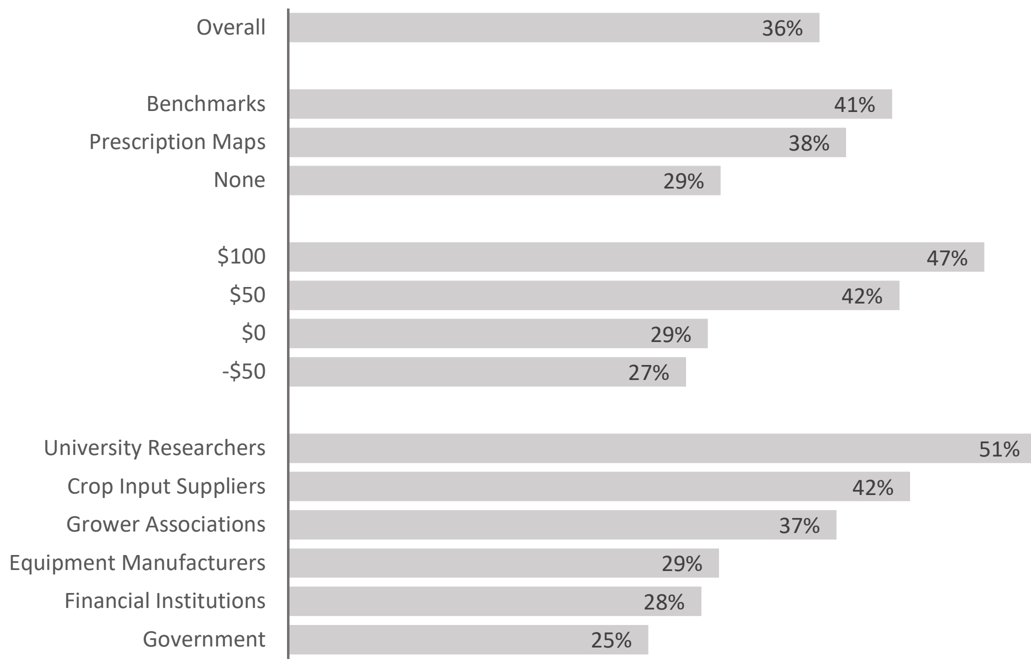
The survey resulted in a panel dataset. Each respondent answered twelve choice questions, which yielded twelve observations for the study. However, not all respondents answered all questions as respondents had the option to refuse to answer. Table 4 shows how the analysis sample was constructed. Observations with missing covariates are removed. This does not bias the results as they are consistent when dummy variables are included for non-response. The final analysis sample includes no missing covariates and has 5265 observations.

Table 4: Analysis sample construction

Starting sample size	6732
Less missing from choice question set	6510
Less missing from technology use question set	6419
Less missing from attitude question set	6245
Less missing income observations	5571
Less other missing covariates	5265

Overall, farmers were willing to participate in a big data program 36% of the time. Figure 4 shows how the big data program participation rate changes when individual attributes of the choice question set are examined. The likelihood of farmer participation in the big data sharing program was highest for university researchers and lowest for a government. Farmers are more willing to share their data in the presence of a non-financial incentive than in the absence of one. In addition, as the financial incentive for participation increased, the percent of farmers that were willing to share their data increased as well. Interestingly, the number of farmers willing to participate when having to pay $50 is almost the same as when there is no financial incentive present. These descriptive statistics are reflected in the results section of this thesis.

Figure 4: Big data program participation rates for select subsamples



METHOD

The responses to the choice questions are analyzed using a heteroscedastic probit, random effects probit, and latent class logit models. I assume the utility the *i*th individual receives from participating in a *j*th big data program is a function of the organization that runs the program, the financial incentive for participation, the non-financial incentive for participation, and a vector of individual attitude characteristics and socio-demographic data. The utility the *i*th individual gains from participating in the *j*th big data program is,

$$u_{i,j} = \sum_{f \epsilon F} \beta_f D_{f,j} + \sum_{g \epsilon G} \beta_g D_{g,j} + \sum_{h \epsilon} \beta_h D_{h,j} + \gamma Z_i + e_{i,j};$$

where,

$F \, \epsilon \, \{Government, University\ researchers, Crop\ input\ suppliers, Grower\ associations, Equipment\ manufacturers, Financial\ institutions\}$,

$D_{f,j} = 1$ if the jth program involves the fth organization,

$D_{f,j} = 0$ otherwise,

β_f represents the effect of the *f*th organization on utility,

$G \, \epsilon \, \{-\$50, \$0, \$50, \$100\}$,

$D_{g,j} = 1$ if the jth program involves the gth financial incentive,

$D_{g,j} = 0$ otherwise,

β_g represents the effect of the *g*th financial incentive on utility,

$H \, \epsilon \, \{None, Benchmark\ statistics, Prescription\ maps\}$,

$D_{h,j} = 1$ if the jth program involves the hth non $-$ financial incentive,

$D_{h,j} = 0$ otherwise,

β_h represents the effect of the *h*th non-financial incentive on utility,

Z_i represents the vector of attitude characteristics and socio-demographic variables associated with individual i,

γ represents the vector of coefficients related to Z_i, and

$e_{i,j}$ represents unobservables.

If one assumes e_i is normally distributed, then the probability that individual i chooses to participate in the *j*th big data program can be represented using the probit link function,

$$\Pr\left(y_{i,j} = 1 | f, g, h, Z_i\right) = \Phi\left(\sum_{f \epsilon F} \beta_f D_{f,j} + \sum_{g \epsilon G} \beta_g D_{g,j} + \sum_{h \epsilon H} \beta_h D_{h,j} + \gamma Z_i\right);$$

where,

$y_{i,j} = 1$ when the *i*th individual chooses to participate in the *j*th big data program, and

$y_{i,j} = 0$ when the *i*th individual chooses to not participate in the *j*th big data program.

This relationship depends on the assumption of independence between observations, and homoscedasticity. However, each individual answered up to twelve choice questions in the survey, and one must account for possible correlation between responses from the same individual. A random effects probit model is chosen because it accounts for this correlation and allows for descriptive statistics that are constant for individual across their responses (such as farm revenue) to be included in the model. The results from a Breusch-Pagan test indicated heteroscedasticity was present in the data, rendering the maximum likelihood estimates for the probit model to be inconsistent. A heteroscedastic probit model was chosen to account for this, with standard errors clustered on the individual to account for correlation between responses from the same individual. The heteroscedastic probit model allows the variance to be impacted by a vector of variables, rather than be fixed at one. The construction of the variance equation is shown in appendix 2. Due to software limitation, there is no model that can fully account for both violations of homoscedasticity and independence, but when shown together some confidence can be placed in the consistency of the results.

Versions of the heteroscedastic probit and random effects probit with different variables included are presented in the results section. In particular, a version including interaction terms between farm size by revenue class and financial incentive is included, as well as a model that formulates the attitude variables through factor analysis. A full explanation of the model construction and can be found in appendix 2.

A latent class logit model is chosen to check for individual heterogeneity in the results. Here, I assume the unobservables in the utility function (e) are logistically distributed. The latent class model performs a partition of the sample into *M* classes. Coefficients are then estimated for each class allowing for variability between classes. The probability that individual *i* is partitioned into class *m* is,

$$Prob(class = m|Z_i) = \frac{\exp\{\theta_m Z_i\}}{\sum_{c=1}^{M} \exp\{\theta_c Z_i\}};$$

where,

Z_i is the set of variables used to partition the sample into classes, and
θ is the vector of coefficients corresponding to variables *Z*.

Within each class, the probability that the individual chooses to participate in the big data program is,

$$\Pr\left(y_{i,j} = 1 | class = m, F_j, G_j, H_j\right) = \frac{\exp\{\beta_{f,m}F_j + \beta_{g,m}G_j + \beta_{h,m}H_j\}}{1 + \exp\{\beta_{f,m}F_j + \beta_{g,m}G_j + \beta_{h,m}H_j\}}.$$

The latent class logit model generates a set of coefficients for partitioning the data (θ), and a set of coefficients for organization ($\beta_{f,m}$), financial incentive ($\beta_{g,m}$), and non-financial incentive ($\beta_{h,m}$) for each class. Individuals are not assigned to a class, but rather they are assigned a probability of being partitioned into each class. θ impacts the probability of being partitioned into a class, while $\beta_{f,m}$, $\beta_{g,m}$, and $\beta_{h,m}$ impact the probability of farmer participation in the big data program. The model is attempting to determine if there are groups of respondents that share preferences distinct from other groups. Variability in coefficients between classes indicates heterogeneity in the results. A full explanation of the latent class analysis can be found in appendix 2.

RESULTS

Table 5 shows the marginal effects for the results from the heteroscedastic probit and random effects probit models. The dependent variable was constructed from the choice questions; it equals one when the farmer choses to participate in the big data sharing program and equals zero when the farmer choses not to participate in the program. Any observation where the farmer chose not to respond was dropped. This does not impact the results, as the models were robust when a dummy variable for non-response was included. Independent variables included in the models are technology use variables, attitude variables, revenue range, and compensation for survey completion. Compensation for survey completion had no statistically significant effect on the probability of program participation for any of the analyses conducted.

1 **Table 5: Effects for heteroscedastic probit and random effects probit models**

	Heteroscedastic Probit		Random Effects Probit	
Log pseudolikelihood	*-2996.27*		*-2550.82*	
Number of observations	*5265*		*5265*	
	Marginal Effect	Std. Err.	**Marginal Effect**	Std. Err.
CHOICE VARIABLES				
Organization (base = government)				
University Researchers	0.268***	0.022	0.341***	0.024
Crop Input Suppliers	0.181***	0.021	0.219***	0.024
Equipment Manufacturers	0.078***	0.019	0.083***	0.021
Grower Associations	0.163***	0.020	0.196***	0.023
Financial Institutions	0.054***	0.019	0.065***	0.022
Non-Financial Incentive (base = none)				
Benchmarks	0.113***	0.014	0.151***	0.016
Prescription Maps	0.076***	0.014	0.108***	0.015
Financial Incentive (base = $0)				
-$50	-0.017	0.017	-0.018	0.018
$50	0.120***	0.019	0.154***	0.020
$100	0.168***	0.019	0.219***	0.022
TECHNOLOGY USE VARIABLES (base= I do not use this technology)				
Yield Monitors				
I use this technology and it improves my farm's performance	0.012	0.039	0.006	0.049
I use this technology but it doesn't improve my farm's performance	-0.025	0.036	-0.034	0.046
GPS Guidance				
I use this technology and it improves my farm's performance	-0.006	0.076	0.026	0.091

	Heteroscedastic Probit		Random Effects Probit	
Log pseudolikelihood	*-2996.27*		*-2550.82*	
Number of observations	*5265*		*5265*	
	Marginal Effect	Std. Err.	**Marginal Effect**	Std. Err.
I use this technology but it doesn't improve my farm's performance	-0.063	0.097	-0.066	0.115
Soil Sampling				
I use this technology and it improves my farm's performance	0.056	0.037	0.074*	0.040
I use this technology but it doesn't improve my farm's performance	0.062	0.055	0.081	0.073
Variable rate technology				
I use this technology and it improves my farm's performance	0.031	0.038	0.019	0.042
I use this technology but it doesn't improve my farm's performance	-0.076	0.054	-0.028	0.073
Automatic Section Control				
I use this technology and it improves my farm's performance	-0.093***	0.029	-0.082**	0.038
I use this technology but it doesn't improve my farm's performance	0.299***	0.098	0.358***	0.125
ATTITUDE VARIABES				
Privacy				
Privacy is important to me	-0.017	0.019	-0.02	0.021
I would be put at a disadvantage if other could access info about my farm	-0.027*	0.016	-0.031*	0.018
I feel comfortable sharing information about my farm	0.059***	0.019	0.075***	0.020
Technology use				
I like to have the latest technology	0.008	0.016	0.013	0.021
I find new technologies easy to use	0.008	0.015	0.002	0.019
New technology is more hassle than it is worth	-0.025	0.018	-0.031*	0.018
I am getting maximum use out of available tech on my farm	0.001	0.019	-0.007	0.018
Farm management				
I have implemented new techniques that have been recommended	0.028	0.021	0.027	0.024
I am proactive in seeking advice	0.017	0.019	0.017	0.024
Precision ag will transform agriculture over the next 20 years	0.013	0.015	0.02	0.018
I know better than others how to manage risk on my farm	-0.032**	0.016	-0.029	0.018
REVENUE RANGE (base = $100,000 to $499,999)				
<$100,000	0.001	0.088	0.008	0.106
$500,000 to $999,999	0.057	0.036	0.068	0.044
$1 million to $2 million	0.053	0.039	0.044	0.051
$2 million to $3 million	0.015	0.066	0.014	0.090
>$3 million	-0.154***	0.050	-0.168***	0.053
COMPENSATION FOR SURVEY COMPLETION (base = $10)				
$20	-0.012	0.032	-0.013	0.040
$30	-0.014	0.039	-0.026	0.046

Significance codes: '' 10%, '**' 5%, '***' 1%*

Organization

Across all model specifications, farmer preferences are consistent for organization. The effects related to organization are interpreted as the percentage point difference in the probability of farmer participation between government running the big data program and the organization of interest running the big data program. All effects are statistically significant at the one percent level.

Farmers are least willing to share their data with government. When the survey was administered (2017), 42 percent of people in the prairie provinces (Alberta, Saskatchewan, and Manitoba) thought the performance of the federal government was very poor. An additional 14.3 percent of people described the federal government's performance as somewhat poor (Mood of Canada Annual Tracking Survey, 2017). The ever-changing nature of government makes measuring long term opinions about it difficult. Farmers' opinions about the current government will inform the results of this study. In addition, farmers could also be less willing to share their data with government out of fear of inciting regulation. Providing information to a body that has the authority to tax and regulate is risky. Some farmers may worry that sharing their data would allow government to catch them for violating current regulations. Alternatively, some farmers may be worried that sharing their data would allow government to uncover hidden trends and enact stricter regulations. Farmers could also be sceptical about what benefits government will be able to generate for the agriculture industry with the data. The government could be slow in innovation development, and disconnected from the problems farmers face.

Second to government, farmers were least willing to share their data with equipment manufacturers and financial institutions. The models found no statistical difference between the effects for these organizations, suggesting farmers are indifferent between them. Both organizations have a direct business relationship with farmers. Farmers could be worried that providing their farm level data to these organizations will affect their pricing strategies to the detriment of farmers. However, crop input suppliers and grower associations also have direct business relationships with farmers, and they are preferred over equipment manufacturers and financial institutions. It could be the case that farmers believe that crop input suppliers and grower associations are better equipped to utilize the collected data to provide long term benefits to farmers than financial institutions and equipment manufacturers. The smallest effect relates to financial institutions. The heteroscedastic (random effects) probit model estimates the effect is 5.4

(6.5) percentage points. In 2016, a five-percentage point change in participation rates corresponded to 13,597 farm operators in Canada (CANSIM table 004-0017). Even a small change in farmer participation rates can have a large impact on the size and success of a big data program.

The models found no statistical difference between the effects for crop input suppliers or for grower associations. These organizations were second only to university researchers in their big data sharing program participation rates. The probability of farmer participation with grower associations in charge was between 16 and 20 percentage points above government, while it was between 18 and 22 percentage points above government with crop input suppliers in charge. Grower associations are not for profit but do have a direct business relationship with farmers. They fund research, create educational programs, and focus on finding new markets and increasing demand for their products. Farmers have a high preference for this group, however it still lags university researchers. Part of this could be a result of farmers' frustration about paying mandatory checkoff levies such as those collected by Sask Pulse. Crop input suppliers could be more preferred because they often provide advice to farmers about which products to use. If they have access to farm level data, the advice they give could be better.

Farmers are most willing to share their data with university researchers. The heteroscedastic probit estimates the likelihood of farmer participation is 26.8 percentage points higher when university researchers run the big data program compared to government, while the random effects probit model estimates the effect is closer to 34.1 percentage points. University researchers could be the most preferred organization because they are not for profit, and do not have a direct business relationship with farmers. Farmers may also want to share their information with university researchers because they support research in agriculture. Research breakthroughs can have a positive effect on farm day to day operation, profit, and environmental quality. Farmers might trust university researchers more because they are viewed as impartial with no political or business agenda.

The ability of different organizations to extract value from a big dataset may also play into farmers willingness to participate. Universities can touch the edges of a multitude of farm problems, and so could potentially extract more value out of a big dataset. This could contribute to universities being the most preferred organization. On the other hand, organizations such as financial institutions have a narrower scope in the services they provide to farmers, and so farmers are less willing to share their data in this case.

Non-financial incentive

The effects for non-financial incentive are statistically and economically significant in the heteroscedastic and random effects probit models. In the heteroscedastic (random effects) probit model, benchmark statistics increase the probability of participation in the big data program by 11.3 (15.1) percentage points, while prescription maps increase the probability of program participation by 7.6 (10.8) percentage points. Benchmarks may be preferred to prescription maps because prescription maps are only useful to those farmers that use variable rate technology (29% of the analysis sample). Benchmark statistics could be useful to every farmer regardless of the equipment available to them. Farmers can use benchmark statistics to compare their own performance with that of comparable farming operations and look for areas where improvements in efficiency could be made. The positive marginal impact on benchmarks could also be explained by the rank income hypothesis. The rank income hypothesis states that people care more about their income rank rather than their absolute level of income (Boyce, Brown, and Moore, 2010). Farmers not only care about their individual productivity, but also are interested in how they perform in relation to other farmers. People gain utility from comparing their level of income to those around them, or to past levels of income they experienced themselves rather than the absolute value of their income (Clark, Frijters, and Shields, 2008).

Financial incentive, revenue range, and interaction terms

Farmers are more willing to participate in a big data program in the presence of a financial incentive. In the heteroscedastic (random effects) probit model, farmers are 12.0 (16.8) percentage points more likely to participate if offered \$50, and 15.4 (21.9) percentage points more likely to participate if offered \$100 than if no incentive is offered. Paying \$50 for the right to participate has no statistically significant effect on farmers' willingness to participate. This is counter to expectations and surprising considering the statistical strength of the coefficients related to positive financial compensation. A positive financial incentive affects participation rates, while a monetary contribution for participation does not.

These financial rewards are relatively small in comparison to total farm revenue, which can range to over \$3 million each year. If farmers have strong privacy preferences these financial rewards should not induce them to share their information so easily. These results are an example of the privacy paradox consistent with Athey, Catalini, and Tucker (2017).

The revenue range \$100,000-\$499,999 was chosen as the base for the farm revenue variables because it is the group with the largest number of farmers in it in the analysis sample. Farm revenue does not affect farmers' willingness to participate in a big data sharing program, except for the largest farms. Farmers whose operations generate the most revenue (>\$3 million per year) are 15.4 (16.8) percentage points less likely to share their data than those that generate \$100,000-\$499,999 in revenue per year in the heteroscedastic (random effects) probit models. Farmers running larger operations generally do not generate off farm income, instead focussing all their effort on managing the farm. Farmers running larger operations have more vested interests in agriculture because they have more at stake financially than farmers running smaller operations. This could cause larger farmers to be more private. Alternatively, larger farmers might not see value beyond the participation incentive for participation. Very large farms generate lots of data, and may have some capacity to do analysis themselves.

A second specification of the models was run including interaction terms between financial incentive and farm revenue. This was done to account for heterogeneity in the results. The effects for financial incentive may not be the same for all respondents. Farmers that generate different levels of revenue may have different marginal utilities for additional income, in this case the financial incentive. The effects for the interaction terms are shown in table 6. The effects for financial incentive remain consistent with what was found in the model excluding control variables.

The overall revenue effects in the model with interaction terms differ slightly from the base model. The effect of >\$3 million becomes insignificant, likely captured by the interaction term between \$100 financial incentive and revenue >\$3 million. This interaction term is negative and statistically significant. In addition, the revenue range \$500,000 to \$999,999 becomes statistically significant and positive, suggesting farmers that generate revenue in this range are more likely to participate in a big data sharing program.

The interaction terms do not reveal underlying heterogeneity in the results. The terms that are statistically significant are the interaction terms between \$100 financial incentive and >\$3 million in revenue, and between \$50 financial incentive and \$500,000 to \$999,999 in revenue (in the random effects probit model only). In the heteroscedastic probit, the Wald test statistic is 19.11 (p-value is 0.22) when testing between the base model and the model including the interaction terms. This suggests that the base model is more appropriate. In the random effects probit, the

Wald test statistic is 23.1 (p-value is 0.08) when testing between the base model and the model including the interaction terms. This suggests the model that includes the interaction terms is more appropriate. There is no overwhelming evidence suggesting systematic underlying heterogeneity exists in the results related to farm revenue and financial incentive.

Table 6: Effects of interaction terms between farm revenue and financial incentive

		Heteroscedastic Probit		Random Effects Probit	
		Marginal Effect	Std. Err.	**Marginal Effect**	Std. Err.
FINANCIAL INCENTIVE (base= $0)					
-$50		-0.032	0.025	-0.036	0.031
$50		0.142***	0.030	0.205***	0.034
$100		0.178***	0.035	0.251***	0.039
REVENUE RANGE (base = $100,000-$499,999)					
<$100,000		0.027	0.117	0.069	0.146
$500,000 to $999,999		0.070*	0.041	0.097**	0.054
$1 million to $2 million		0.040	0.049	0.045	0.059
$2 million to $3 million		-0.011	0.076	0.029	0.099
>$3 million		-0.089	0.074	-0.092	0.082
INTERACTION TERMS					
Fin. Incentive	**Revenue Range (base = $100,000-$499,999)**				
-$50	<$100,000	-0.186	0.240	-0.145	0.133
	$500,000 to $999,999	0.018	0.041	0.013	0.051
	$1 million to $2 million	0.053	0.049	0.069	0.057
	$2 million to $3 million	0.207	0.198	0.119	0.076
	>$3 million	0.024	0.116	-0.011	0.127
$50	<$100,000	-0.005	0.161	-0.057	0.162
	$500,000 to $999,999	-0.052	0.039	-0.088**	0.048
	$1 million to $2 million	0.006	0.068	-0.031	0.052
	$2 million to $3 million	-0.016	0.213	-0.108	0.085
	>$3 million	-0.106	0.099	-0.153	0.122
$100	<$100,000	-0.026	0.154	-0.061	0.133
	$500,000 to $999,999	-0.007	0.054	-0.033	0.053
	$1 million to $2 million	-0.006	0.072	-0.033	0.055
	$2 million to $3 million	0.179	0.306	-0.054	0.094
	>$3 million	-0.176**	0.084	-0.198**	0.106

Significance codes: '' 10%, '**' 5%, '***' 1%*

Technology use

None of the technology use questions yielded a statistically significant result except the use of automatic section control. In the results of the survey, only four individuals (totalling 48 observations) answered "I use this technology, but it doesn't improve my farm's performance" for automatic section control. The rest of the respondents were equally split between not using the technology and answering, "I use this technology and it improves my farm's performance". The limited response to "I use this technology, but it doesn't improve my farm's performance" reduces confidence in the estimated effects. Forty-eight observations over four individuals is too small of a sample to draw a decisive conclusion. However, more confidence can be place in the marginal effect related to "I use this technology and it improves my farm's performance". Those farmers that responded positively to that statement are between 8.2 and 9.5 percentage points less likely to participate in a big data program than those that do not use automatic section control.

Attitude Variables

The attitude variables are analyzed in the subsets of statements on privacy (first three statements), technology use (next four statements), and farm management (final four statements). The statements are first analyzed individually, and then using factor analysis. The marginal effects for the individual analysis are interpreted as the percentage point change in willingness to participate in a big data program from a one-point increase in the level of agreement with the statement. The level of agreement is measured on a five-point scale, with one being strongly disagree and five being strongly agree. The marginal effects for the individual analysis can be found in table 1, while the marginal effects for the factor analysis can be found in table 7.

Attitudes concerning privacy had the largest impact on farmers' willingness to participate in a big data program. The only statement from any category that showed consistently strong statistically significant results was "I feel comfortable sharing information about my farm". This is intuitive, as the choice questions are attempting to capture willingness to share information. It lends evidence that survey respondents were consistent in their preferences when asked about privacy. The marginal effect on "I would be put at a disadvantage if other could access info about my farm" was weakly significant and negative.

None of the technology use statements were strongly statistically significant, however, "New technology is more hassle than it is worth" yielded a weakly significant negative result in the random effects probit model. Those that struggle to see the value of new technology are less willing to adopt new technology and farming practices, including participation in a big data program. In addition, only one of the farm management statements was found to be weakly significant in the heteroscedastic probit model. “I know better than others how to manage risk on my farm” was found to have a negative relationship with farmers’ willingness to participate in a big data program. These results suggest that as a farmer has a more independent managing style, they become less willing to participate in a big data program. The low statistical significance of most of the effects of the attitude variables could be a result of multicollinearity in the data. The statements are each attempting to capture one of three measures (privacy attitudes, technology use attitudes, or farm management attitudes) and so should be highly correlated. Tests for joint significance within each subcategory of the attitude variables in the heteroscedastic (random effects) probit revealed p-values of 0.00 (0.00) for privacy, 0.51 (0.26) for technology use, and 0.03 (0.24) for farm management.

Factor analysis condenses the statements in each category into one measure, capturing as much original variation in the statements as possible. A more thorough explanation of the construction of the factors can be found in appendix 2. The marginal effects are shown in table 3. The sizes of the marginal effects are uninterpretable as the exact construction and units of each factor is unknown. However, I multiply the marginal effect by the standard deviation for each factor to gain some insight into the economic significance of the effects.

Table 7: Effects for factors generated by factor analysis

	Het Probit		RE Probit	
	Marginal Effect	Std. Err.	**Marginal Effect**	Std. Err.
Privacy (increases with increasing strictness in privacy preferences)	-0.055***	0.010	-0.076***	0.012
Technology use (increases with more favourable attitudes towards technology)	0.021	0.014	0.021	0.017
Farm management (increases with increasing progressiveness in farm management style)	0.025*	0.013	0.025	0.016

Significance codes: ‘’ 10%, ‘**’ 5%, ‘***’ 1%*

Privacy has the largest impact on farmers’ willingness to participate, and the only strongly statistically significant one. The standard deviation related to the privacy factor is 1.29. An increase of one standard deviation in the privacy factor relates to a decrease of 7.1 (9.8) percentage points in the probability of farmer participation in a big data sharing program in the heteroscedastic (random effects) probit model. As privacy attitudes become stricter, willingness to participate in a big data program is decreased. Technology use attitudes and farm management attitudes do not significantly impact farmers' willingness to participate in a big data sharing program, although farm management attitudes have a weak relationship with program participation in the heteroscedastic probit model. The results from the factor analysis are consistent with the results from the individual analysis.

Robustness Check: How well does the analysis sample represent the overall population?

To test for self-selection bias, I compare respondents in the first and fourth quartiles of respondents by the date the survey was completed. Respondents completed the survey voluntarily, meaning there were farmers asked to complete the survey but that chose not to. There could be self-selection bias if those that chose to respond to the survey are different than those that chose not to respond. Out of farmers that chose to respond, there were fast responders (those in the first quartile of respondents) and slow responders (those in the last quartile of respondents). Farmers that did not respond to the survey are more likely to be similar to slow respondents than fast respondents. Assuming non-respondents are similar to slow respondents, differences between slow respondents and fast respondents lends evidence towards a sample self-selection bias.

To test this, I compare the first and fourth quartiles of respondents in order of response time by running a pooled and segmented heteroscedstic probit and random effects probit excluding observations in the second and third quartiles. I use a Wald test to determine if the pooled model is more appropriate than the segmented model. The results are in table 8. There is no evidence of sample selection bias in the heteroscedastic probit model, however there is evidence of sample selection bias in the random effects probit model. The stark differences in the results from the Wald tests can be attributed to the marginal effect on financial institutions. In the random effects probit model, this marginal impact is statistically significant in quartile 1 but not in quartile 4. This is the only significant difference found between quartiles. The ordering of organization, financial incentive, and non-financial incentive remains consistent between quartiles. The full result tables from the quartile analysis can be found in appendix 1.

Table 8: Wald test results for quartile analysis

	Wald Test Statistic	P-value	Conclusion
Het. Probit	2.28	1.00	There is no statistical difference between any coefficients for the 1st and 4th quartiles. No evidence of sample selection bias.
RE Probit	30.80	0.07	At a 10% level of significance, at least one coefficient is different between the 1st and 4th quartiles. Evidence of sample selection bias.

In addition to testing between models by quartiles, I performed a comparison of means for the attitude variables between quartiles. The results are found in table 9. A low p-value indicates significant differences between the means in quartile 1 and quartile 4. The respondents in quartile 4 appear to be more technologically savvy than those in quartile 1. However, the differences do not appear to be large, and the responses between statements appears to be consistent (for example, respondents agreed more with the first statement than with the second statement). In addition, those in quartile 4 are statistically significantly younger by 1.73 years.

Table 9: Comparison of means between Q1 and Q4 for attitude variables

	Q1	Q4	p-value[1]
Privacy is important to me	4.02	4.06	0.39
I would be put at a disadvantage if other could access info about my farm	3.06	3.27	0
I feel comfortable sharing information about my farm	3.22	3.33	0
I like to have the latest technology	3.22	3.45	0
I find new technologies easy to use	3.24	3.43	0
New technology is more hassle than it is worth	2.69	2.55	0
I am getting maximum use out of available tech on my farm	3.32	3.28	0.28
I have implemented new techniques that have been recommended	3.64	3.72	0.01
I am proactive in seeking advice	3.9	3.86	0.36
Precision ag will transform agriculture over the next 20 years	4.04	3.9	0
I know better than others how to manage risk on my farm	3.71	3.64	0.06

[1] *p-value from a paired t-test comparing group means*

Latent Class Analysis

Latent class analysis was performed to determine if there are heterogeneities in the population of grain farmers in Saskatchewan. Two sets of coefficients are generated in latent class analysis. The first varies between classes, in this analysis, organization, financial incentive, and non-financial incentive. The second set of coefficients relates a set of variables (*Z*) to the probability of class membership. The probability that individual i is in class m is:

$$Prob(class = m|Z_i) = \frac{\exp\{\theta_m Z_i\}}{\sum_{c=1}^{M} \exp\{\theta_c Z_i\}},$$

where,

Z_i is the set of variables used to partition the sample into classes, and
θ is the vector of coefficients corresponding to variables *Z*.

The latent class analysis is performed on four different sets of variables in *Z*. The first includes an intercept term, the second includes socio-demographic indicators, the third includes the attitude statements, and the fourth includes technology use information. The results presented here are from the model including socio-demographic indicators (Z_{2i}),

(1) $Z_{1i} = 1,$
(2) $Z_{2i} = 1 + age_i + gender_i + \sum_{e\epsilon E} education_{ei} + \sum_{r\epsilon R} revenue_{ri},$
(3) $Z_{3i} = 1 + \sum_{a\epsilon A} attitude\ statements_{ai},$
(4) $Z_{4i} = 1 + \sum_{t\epsilon T} I\ use\ this\ tech.but\ it\ doesn't\ improve\ my\ farms\ performance_{ti} + \sum_{t\epsilon T} I\ use\ this\ tech\ and\ it\ improves\ my\ farms\ performance_{ti},$

where,

i is an index for individual,
$E\ \epsilon\ \{Less\ than\ highchool, Some\ post\ secondary, College\ degree, University\ degree, Graduate\ degree\},$
$R\ \epsilon\ \{< \$100{,}000, \$500{,}000 - \$999{,}999, \$1M - \$2M, \$2M - \$3M, > \$3M\},$
$A\ \epsilon\ \{11\ attitude\ statements\}$, and
$T\ \epsilon\ \{Yield\ monitor, GPS, Soil\ sampling, Variable\ rate\ technology, Automatic\ section\ contol\}.$

Using the Bayesian Information Criterion (BIC), a three-class model was selected. This is the most appropriate model for all Z_i, as the BIC is lowest (table 10).

Table 10: BIC values for different numbers of classes and *Z*

Number of Classes	Z_{1i}	Z_{2i}	Z_{3i}	Z_{4i}
2	5570	5648	5608	5626
3	**5476**	**5646**	**5592**	**5609**
4	5498	*	5700	5693
5	5543	*	*	5816

** model convergence was not achieved*

The coefficients for the three-class model with socio-demographic indicators in the membership equation are shown in table 11. Class one contained 33% of respondents, class two contained 20% of respondents, and class three contained 47% of respondents. Based on the generated coefficients, class one is labelled as “Likely Non-participants”, class two is labelled as “Money Lovers”, and class three is labelled as “Majority Group”. Likely Non-participants get their name from the large negative intercept in class one, Money Lovers get their name from the large coefficients on financial incentive, and the Majority Group gets it’s name because 47% of respondents fall into this category.

Table 11: Coefficients and standard errors for three class model, socio-demographics in the membership equation (Z_{2i})

	Class1		Class 2		Class 3	
Class probabilities	*33%*		*20%*		*47%*	
	Coef.	Std. Err.	**Coef.**	Std. Err.	**Coef.**	Std. Err.
Intercept	-6.05***	0.941	-1.05***	0.302	-2.32***	0.218
Organization (base = government)						
University Researchers	2.63***	0.828	1.51***	0.342	2.04***	0.184
Crop Input Suppliers	2.41***	0.866	1.60***	0.336	1.19***	0.185
Equipment Manufacturers	0.333	1.29	1.07***	0.323	0.471**	0.196
Grower Associations	1.97**	0.871	1.31***	0.331	1.17***	0.180
Financial Institutions	1.18	0.931	0.748**	0.316	0.342*	0.187
Non-Financial Incentive (base = none)						
Benchmarks	0.701*	0.390	0.770***	0.234	1.01***	0.128
Prescription Maps	0.954**	0.380	0.466**	0.217	0.719***	0.125
Financial Incentive (base = $0)						
-$50	0.099	0.515	0.310	0.229	0.007	0.148
$50	0.933**	0.448	2.76***	0.663	0.607***	0.156
$100	1.21***	0.447	2.06***	0.332	1.13***	0.152

Significance codes: ‘’ 10%, ‘**’ 5%, ‘***’ 1%*
Number of observations: 5265
Log-likelihood: -2570.14

The estimated coefficients change between classes, but the ordering of preferences remain consistent except for small variations. The orderings of preferences for organization are consistent with previous results, except crop input suppliers are the most preferred organization in class two rather than university researchers, and the coefficients on equipment manufacturers and financial institutions are not significantly different from zero in class one.

In previous results, benchmark statistics were preferred to prescription maps for non-financial incentive. This is true in class two and class three, but the opposite is true in class one. Prescription maps are valuable only to the subset of farmers that use variable rate technology. To these farmers, the value of prescription maps may be high, while other farmers may not place much value on prescription maps. However, when analyzing the membership equation with technology use variables included (Z_{4i}), the use of variable rate technology is not related to class membership. Individuals are not more likely to be in class one if they use variable rate technology.

Consistent with previous results in this thesis, a monetary contribution from the farmer for participation does not decrease farmers' willingness to participate in a big data program in any class. Coefficients on positive financial incentive are consistent with previous results in class one and class three, but in the Money Lovers class a \$50 incentive is proffered to a \$100 incentive. In addition, the coefficients in class two are much larger than those in class one or class three, suggesting individuals in class two are more easily incentivized by monetary measures than other people.

The largest difference between classes emerges in the intercept term. People in class one are least likely to participate in a big data program, followed by people in class two, and people in class three. The size of the intercept in class one is 6.4 times the size of the intercept in class three, and 2.8 times the size of the intercept in class two. Identifying which class an individual belongs to is important in understanding their willingness to participate in a big data program.

The overall partial effects are shown in table 2. They are consistent with the effects found in previous results.

Table 12: Overall partial effects for latent class logit, socio-demographic indicators in the membership equation (Z_{2i})

	Partial Effect	Elasticity
Organization (base = government)		
University Researchers	0.413***	0.262
Crop Input Suppliers	0.324***	0.206
Equipment Manufacturers	0.106	0.067
Grower Associations	0.283***	0.180
Financial Institutions	0.136**	0.086
Non-Financial Incentive (base = none)		
Benchmarks	0.167***	0.212
Prescription Maps	0.145***	0.183
Financial Incentive (base = $0)		
-$50	0.019	0.018
$50	0.221***	0.210
$100	0.260***	0.248

Significance codes: '' 10%, '**' 5%, '***' 1%*
Number of observations: 5265

None of the vectors included in the membership equations (Z) yielded results that had significant impacts on the probability of class membership. Full tables showing these results can be found in appendix 1. The coefficients with socio-demographic indicators in the membership equation are shown in table 13.

Table 13: Coefficients impacting class membership (base = class 3), socio-demographic indicators included in the membership equation (Z_{2i})

	Class 1		Class 2	
	Coef.	Std. Err.	**Coef.**	Std. Err.
Intercept	-0.203	0.904	-0.629	1.21
Age	0.007	0.012	0.008	0.015
Sex	0.247	0.600	-0.125	0.624
Education (base = highschool)				
Less than highschool	-0.537	0.592	-0.133	0.725
Some post secondary education	-0.913**	0.365	-0.545	0.472
College degree	0.047	0.493	0.263	0.600
University degree	-1.11***	0.427	-0.563	0.527
Graduate degree	-0.504	0.725	-0.690	1.00
Revenue range (base = $100,000 - $499,999)				
<$100,000	0.761	0.788	0.615	0.911
$500,000 to $999,999	-0.689**	0.313	-0.262	0.334
$1 million to $2 million	-0.481	28.6	-0.451	30.2
$2 million to $3 million	-0.420	98.4	-0.841	96.8
>$3 million	0.968	25.9	-2.05	32.1

Significance codes: '' 10%, '**' 5%, '***' 1%*
Number of observations: 5265

The latent class analysis indicates that some heterogeneities exist within the farmer population in Saskatchewan. However, those heterogeneities are not strongly related to any of the Z vectors specified in this thesis. It is therefore difficult to predict which class an individual belongs to. Without the ability to identify which class an individual belongs to; the latent class analysis has few real world implications. The classes may exist, but if we cannot match them with the people that are in them, we cannot tailor incentives to classes.

DISCUSSION AND CONCLUSION

The development of a big data sharing program is the first step for researchers to perform regional and Canada wide big data analysis in agriculture. Government, farmers, producer groups, and corporations are currently missing out on the value of potential technological and organizational advancements made doing this type of analysis. The construction of a big data sharing program brings these abstract benefits closer to reality. The value of a big database increases with size, so high farmer participation rates are important. This thesis examines farmers' willingness to participate in a big data program, and how the organization administering the program, financial incentive for participation, and non-financial incentives for participation impact farmers' willingness to participate.

The results are consistent across all models. Farmers are most willing to share their data with university researchers, then crop input suppliers or grower associations, followed by financial institutions or equipment manufacturers, and government. Farmers are more willing to share their data in the presence of any positive financial or non-financial incentive, but a negative financial incentive has no impact on willingness to participate.

It is important to note that the construction of the survey doesn't allow for the big data sharing program to vary in its structure between organizations. In reality, a big data program run by a university could look very different than one run by a different organization, both in terms of how it is administered, and what the end use goals are. Farmers may be more sympathetic with the objectives of a certain organization, and this could contribute the differences in willingness to participate between organizations. In addition, the farmers surveyed were commercial farmers rather than hobby farmers, which could also impact willingness to participate.

These results imply that a big data program run by a university would generate higher response rates than one run by the other types of organizations studied in this thesis. Organizations other than universities interested in pioneering a big data sharing program may be better off contracting through a university to run it. This is especially true for government, as the difference in participation rates could be as high as 34 percentage points. For organizations such as crop input suppliers and grower associations, the difference in participation rates ranges from 8.7 to 14.5 percentage points, so the additional cost of contracting through a university may not be worth it. These companies could make up for the lower participation rates by increasing the financial incentive for participation. This would also increase the cost of the program, but it could still be more affordable than paying a university to administer it. However, contracting through a university must be done carefully. If it becomes blatantly obvious that the university is simply acting as a middle man, trust in the university could be eroded, and participation rates could drop.

A larger financial compensation increases the participation rate but also increases cost with each respondent. The presence of a non-financial incentive also increases the participation rate, but costs do not increase linearly with the number of respondents. Non-financial incentives can be more difficult to offer because they must be tailored for each respondent. Although for some organizations such as grower associations, providing these non-financial benefits to farmers may be the very reason for the big data program in the first place. Program administrators can make a fixed investment cost in the development of a program that generates non-financial incentives for respondents quickly and easily. The least cost method of increasing participation rates will depend on the efficiency of generating non-financial incentives, and the number of farmers participating in the big data program.

The latent class logit model revealed some heterogeneities in the farmer population. The ordering of coefficients remained relatively consistent between classes, but large differences in the intercept term were found. Farmers in class one are least likely to participate in a big data program, while farmers in class three are most likely to participate in a big data program. The explanatory power of socio-demographic indicators, attitude statements, and current technology use on class membership was low. These results do not find a good indicator of class membership.

Farmers may be hesitant to share their data if they feel someone else is profiting off it. There could be a sense that the data belongs to the farmer, and any profits made from it should also go to the farmer. This could be the case for big data programs operated by for profit groups like crop input suppliers, equipment manufacturers, and financial institutions. Grower associations and universities are generally non-profit. Farmers may prefer these because it does not feel as if someone is getting rich off their backs.

There are also positive impacts on farmers for participation. It is in farmers' best interest to foster innovation in agriculture, as they are the ones that will see the eventual impacts. Innovations developed by university researchers could have significant positive impacts on farmers. Crop input suppliers, equipment manufacturers, and financial institutions may be better able to serve farmers if they have access to more detailed information about them. This could be a good thing for both farmers and the companies serving them.

A limitation of this study is that the entire sample was induced to respond to the survey through financial measures. This suggests that they are a group that cares about marginal financial rewards. Farmers who place less value on money (or a higher cost on responding to surveys) might be excluded from the sample. People incentivized by money once may be more inclined to be incentivized by it again. This could inflate the effects of the financial compensation variables. In addition, the survey was administered online. Farmers must have had an email address, and access to the internet to complete the survey. This might have excluded less technologically oriented individuals from the sample. The quartile analysis suggests that sample selection bias is not a problem, however there could be biases that the quartile analysis was not able to detect.

Questions surrounding data ownership remain. Depending on the legal ownership of the data, the structure of a big data sharing program could vary widely. If farmers do not have the legal right to their data, corporations may be able to collect it without farmers permission. This is an area in which government must clarify the rules before a program is put in place.

Further research could be done exploring the latent class model. Identifying indicators that are related to class membership could open some real world implication for the latent class logit model. If there is an observable characteristic with which the probability of class membership is highly correlated, identifying which farmers are predisposed to more readily sharing their data could help match incentives with individuals. Perhaps there is a system where some individuals are offered one incentive, and other individuals are offered a different incentive. This could reduce costs for the program administrator.

REFERENCES

Athey, S., Catalini, C., and Tucker, C. 2017. The Digital Privacy Paradox: Small Money, Small Costs, Small Talk. National Bureau of Economic Research Working Paper.

Barnes, S. 2006. A privacy paradox: Social networking in the United States. *First Monday, 11*(9).

Bloomberg. 2018. Facebook Cambridge Analytica Scandal: 10 Questions Answered. *Fortune.* http://fortune.com/2018/04/10/facebook-cambridge-analytica-what-happened/.

Boyce, C., Brown, G., and Moore, S. 2010. Money and Happiness: Rank of Income, Not Income, Affects Life Satisfaction. *Psychological Science,* 21(4):471-475.

Boyer, A., Engleking, E., and Gudas, S. 2015. Farmer perceptions of big data in agriculture. *Journal of Purdue Undergraduate Research*, 5: 82-83.

Bronson, K., and Knezevic, I. 2016. Big Data in food and agriculture. *Sage Journals*.

Clark, A., Frijters, P., and Shields, M. 2008. Relative Income, Happiness, and Utility: An Explanation for the Easterlin Paradox and Other Puzzles. *Journal of Economic Literature,* 46(1):95-144.

Coble, K. 2018. Risk Management in an Era of Big Data and Changing Climate. *Presented at the Eighth Annual Canadian Agri-Food Policy Conference, Chateau Laurier, Ottawa ON.*

Cox, M., and Ellsworth, D. 1997. Application-controlled demand paging for out-of-core visualization. Proceedings of the 8th conference on Visualization '97. IEEE Computer Society Press, Los Alamitos, CA, USA.

Cukier, K., and Mayer-Schoenberger, V. 2013. The rise of big data: How it's changing the way we think about the world. *Foreign Affairs*, 92(3): 28-[ii].

Daberkow, S., and McBride, W. 2003. Farm and Operator Characteristics Affecting the Awareness and Adoption of Precision Agriculture Technologies in the US. *Precision Agriculture,* 4:163-177.

Debatin, B., Lovejoy, J.P., Horn, A.K., & Hughes, B.N. 2009. Facebook and Online Privacy: Attitudes, Behaviours, and Unintended Consequences. *Journal of Computer-Mediated Communications*, 15: 83-108.

Fan, J., Han, F., and Liu, H. 2014. Challenges of Big Data analysis. *National Science Review,* 1:293–314.

George, G., Haas, M., and Pentland, A. 2014. Big Data and Management. *Academy of Management Journal*, 57(2): 321-326.

Katz, M., and Shapiro, C. 1985. Network externalities, competition, and compatibility. *American Economic Review,* 75: 424.

Katz, M., and Shapiro, C. 1986. Technology adoption in the presence of network externalities. *Journal of Political Economy*, 94:822.

Kamilaris, A., Kartakoullis, A., and Prenafeta-Boldú, F. 2017. A review on the practice of big data analysis in agriculture. *Computers and Electronics in Agriculture,* 143:23-37.

Larson, J., Roberts, R., English, B., Larkin, S., Marra, M., Martin, S., Paxton, K., and Reeves, J. 2008. Factors affecting farmer adoption of remotely sensed imagery for precision management in cotton production. *Precision Agriculture,* 9:195-208.

Lyseng, R. 2018. U.S. farmers fight for right to repair. *Western Producer, March 8, 2018.*

Miller, Griffin, Bergtold, Ciampitti, and Sharda. 2017. Farmers' Adoption Path of Precision Agriculture Technology. *Advances in Animal Biosciences,* 8(2):708-712.

Mood of Canada Annual Tracking Survey. (2017). *Nanos.* http://www.nanos.co/dataportal/.

Norberg, P., Horne, D., and Horne, D. 2007. The Privacy Paradox: Personal Information Disclosure Intentions versus Behaviours. *Journal of Consumer Affairs,* 41(1):100-126.

Olson, J., Grudin, J., and Horvitz, E. 2005. A study of preferences for sharing and privacy. *CHI '05 Extended Abstracts on Human Factors in Computing Systems (CHI EA '05). ACM, New York, NY, USA,* 1985-1988.

Pavolotsky, J. 2013. Privacy in the Age of Big Data. *The Business Lawyer*, 69(1): 217-225.

Pierce, F., and Nowak, P. 1999. Aspects of Precision Agriculture. *Advances in Agronomy,* 67:1-85.

Poppe, K., Wolfert, J., Verdouw, C., and Renwick, A. 2015. A European perspective on the economics of Big Data. *Farm Policy Journal* 12.1: 11-19.

Schemper, J. 2014. Efficiency of Combine Usage: a Study of Combine Data Comparing Operators and Combines to Maximize Efficiency. *A thesis submitted in partial fufillment of the requirements for the degree Master of Agribusiness, Department of Agricultural Economics, Kansas State University, Manhattan, Kansas.*

Sonka, S. 2014. Big data and the ag sector: More than lots of numbers. *International Food and Agribusiness Management Review,* 17(1): 1-20.

Sonka, S. 2016. Big data: fueling the next evolution of agricultural innovation. *Journal of Innovation Management,* 4(1):114.

Sonka, S., and Coaldrake, K. 1996. Cyberfarm: What does it look like? What does it mean? *American Journal of Agricultural Economics*, 78(5): 1263-1268.

Statistics Canada. *Table 004-0006 - Census of Agriculture, farms classified by total gross farm receipts at 2015 constant dollars, Canada and provinces, every 5 years,* CANSIM. Last updated Dec. 8, 2017.

Statistics Canada. *Table 004-0017- Census of Agriculture, number of farm operators by sex, age and paid non-farm work, Canada and provinces, every 5 years,* CANSIM. Last updated Dec. 8, 2017.

Statistics Canada. *Table 004-0110 - Socioeconomic overview of the farm population, number of operators per farm by occupation and level of educational attainment, every 5 years,* CANSIM. Last updated Nov. 26, 2013.

Sun, S., Cegielski, C., Jia, L., and Hall, D. 2018. Understanding the Factors Affecting the Organizational Adoption of Big Data. *Journal of Computer Information Systems,* 58(3):193-203.

Sykuta, M. 2016. Big Data in Agriculture: Property Rights, Privacy and Competition in Ag Data Services. *The International Food and Agribusiness Management Review,* 19(A).

Tamirat, T., Pedersen, S., and Lind, K. 2018. Farm and operator characteristics affecting adoption of precision agriculture in Denmark and Germany. *Acta Agriculturae Scandinavica, Section B — Soil & Plant Science,* 68(4): 349-357.

Trujillo, G., Kim, C., Jones, S., Garcia, R., and Murray, J. 2015. Virtualizing Hadoop: How to Install, Deploy, and Optimize Hadoop in a Virtualized Architecture. *VMware Press*.

Wiens, K. 2015. We can't let John Deere destroy the very idea of ownership. *Wired, April 21, 2015.*

Wolfert, S., Ge, L., Verdouw, C., and Bogaardt, M. 2017. Big Data in Smart Farming- A review. *Agricultural Systems,* 153:69-80.

APPENDIX 1: Full Results Tables

Table A1.1: Specification 1: Control Variables

	Heteroscedastic Probit		Random Effects Probit	
Log pseudolikelihood	*-2996.27*		*-2550.82*	
Number of observations	*5265*		*5265*	
	Marginal Effect	Std. Err.	**Marginal Effect**	Std. Err.
CHOICE VARIABLES				
Organization (base = government)				
University Researchers	0.268***	0.022	0.341***	0.024
Crop Input Suppliers	0.181***	0.021	0.219***	0.024
Equipment Manufacturers	0.078***	0.019	0.083***	0.021
Grower Associations	0.163***	0.020	0.196***	0.023
Financial Institutions	0.054***	0.019	0.065***	0.022
Non-Financial Incentive (base = none)				
Benchmarks	0.113***	0.014	0.151***	0.016
Prescription Maps	0.076***	0.014	0.108***	0.015
Financial Incentive (base = $0)				
-$50	-0.017	0.017	-0.018	0.018
$50	0.120***	0.019	0.154***	0.020
$100	0.168***	0.019	0.219***	0.022
TECHNOLOGY USE VARIABLES (base= I do not use this technology)				
Yield Monitors				
I use this technology and it improves my farm's performance	0.012	0.039	0.006	0.049
I use this technology but it doesn't improve my farm's performance	-0.025	0.036	-0.034	0.046
GPS Guidance				
I use this technology and it improves my farm's performance	-0.006	0.076	0.026	0.091
I use this technology but it doesn't improve my farm's performance	-0.063	0.097	-0.066	0.115
Soil Sampling				
I use this technology and it improves my farm's performance	0.056	0.037	0.074*	0.040
I use this technology but it doesn't improve my farm's performance	0.062	0.055	0.081	0.073
Variable rate technology				
I use this technology and it improves my farm's performance	0.031	0.038	0.019	0.042
I use this technology but it doesn't improve my farm's performance	-0.076	0.054	-0.028	0.073
Automatic Section Control				
I use this technology and it improves my farm's performance	-0.093***	0.029	-0.082**	0.038
I use this technology but it doesn't improve my farm's performance	0.299***	0.098	0.358***	0.125

	Heteroscedastic Probit		Random Effects Probit	
Log pseudolikelihood	*-2996.27*		*-2550.82*	
Number of observations	*5265*		*5265*	
	Marginal Effect	Std. Err.	**Marginal Effect**	Std. Err.
ATTITUDE VARIABES				
Privacy				
Privacy is important to me	-0.017	0.019	-0.02	0.021
I would be put at a disadvantage if other could access info about my farm	-0.027*	0.016	-0.031*	0.018
I feel comfortable sharing information about my farm	0.059***	0.019	0.075***	0.020
Technology use				
I like to have the latest technology	0.008	0.016	0.013	0.021
I find new technologies easy to use	0.008	0.015	0.002	0.019
New technology is more hassle than it is worth	-0.025	0.018	-0.031*	0.018
I am getting maximum use out of available tech on my farm	0.001	0.019	-0.007	0.018
Farm management				
I have implemented new techniques that have been recommended	0.028	0.021	0.027	0.024
I am proactive in seeking advice	0.017	0.019	0.017	0.024
Precision ag will transform agriculture over the next 20 years	0.013	0.015	0.02	0.018
I know better than others how to manage risk on my farm	-0.032**	0.016	-0.029	0.018
REVENUE RANGE (base = $100,000 to $499,999)				
<$100,000	0.001	0.088	0.008	0.106
$500,000 to $999,999	0.057	0.036	0.068	0.044
$1 million to $2 million	0.053	0.039	0.044	0.051
$2 million to $3 million	0.015	0.066	0.014	0.090
>$3 million	-0.154***	0.050	-0.168***	0.053
COMPENSATION FOR SURVEY COMPLETION (base = $10)				
$20	-0.012	0.032	-0.013	0.040
$30	-0.014	0.039	-0.026	0.046

Significance codes: '' 10%, '**' 5%, '***' 1%*

Table A1.2: Variables and coefficients impacting variance for base model heteroscedastic probit

	Coef.	Std. Err
CHOICE VARIABLES		
Organization (base = government)		
University Researchers	-0.374	0.248
Crop Input Suppliers	-0.409*	0.214
Equipment Manufacturers	-0.253	0.195
Grower Associations	-0.420**	0.199
Financial Institutions	-0.174	0.218
Non-Financial Incentive (base = none)		
Benchmarks	0.008	0.093
Prescription Maps	0.258**	0.104
Financial Incentive (base = $0)		
-$50	0.085	0.138
$50	0.009	0.144
$100	0.149	0.142
ATTITUDE STATEMENTS		
Privacy is important to me	0.001	0.167
I would be put at a disadvantage if other could access info about my farm	-0.075	0.084
I feel comfortable sharing information about my farm	0.013	0.135
I know better than others how to manage risk on my farm	-0.079	0.143
REVENUE RANGE (base = $100,000 to $499,999)		
<$100,000	0.361	0.622
$500,000 to $999,999	0.05	0.214
$1 million to $2 million	0.211	0.279
$2 million to $3 million	1.34**	0.671
>$3 million	0.094	0.421
COMPENSATION FOR SURVEY COMPLETION (base = $10)		
$20	0.191	0.217
$30	0.415	0.293

Table A1.3: Specification 2: Interaction terms between financial incentive and revenue range

	Heteroscedastic Probit		Random Effects Probit	
Log pseudolikelihood	*-2988.39*		*-2532.71*	
Number of observations	*5265*		*5265*	
	Marginal Effect	Std. Err.	**Marginal Effect**	Std. Err.
CHOICE VARIABLES				
Organization (base = government)				
University Researchers	0.269***	0.022	0.341***	0.024
Crop Input Suppliers	0.180***	0.021	0.217***	0.023
Equipment Manufacturers	0.077***	0.019	0.084***	0.021
Grower Associations	0.162***	0.021	0.195***	0.023
Financial Institutions	0.054***	0.020	0.065***	0.022
Non-Financial Incentive (base = none)				
Benchmarks	0.113***	0.014	0.150***	0.016
Prescription Maps	0.076***	0.014	0.107***	0.015
Financial Incentive (base = $0)				
-$50	-0.032	0.025	-0.036	0.031
$50	0.142***	0.030	0.205***	0.034
$100	0.178***	0.035	0.251***	0.039
TECHNOLOGY USE VARIABLES (base= I do not use this technology)				
Yield Monitors				
I use this technology and it improves my farm's performance	0.014	0.039	0.005	0.049
I use this technology but it doesn't improve my farm's performance	-0.023	0.037	-0.036	0.046
GPS Guidance				
I use this technology and it improves my farm's performance	-0.004	0.084	0.026	0.091
I use this technology but it doesn't improve my farm's performance	-0.069	0.103	-0.067	0.115
Soil Sampling				
I use this technology and it improves my farm's performance	0.055	0.038	0.076*	0.040
I use this technology but it doesn't improve my farm's performance	0.065	0.056	0.083	0.073
Variable rate technology				
I use this technology and it improves my farm's performance	0.027	0.039	0.018	0.042
I use this technology but it doesn't improve my farm's performance	-0.069	0.057	-0.029	0.073
Automatic Section Control				
I use this technology and it improves my farm's performance	-0.095***	0.030	-0.082**	0.039
I use this technology but it doesn't improve my farm's performance	0.292***	0.100	0.360***	0.126

	Heteroscedastic Probit		Random Effects Probit	
Log pseudolikelihood	*-2988.39*		*-2532.71*	
Number of observations	*5265*		*5265*	
	Marginal Effect	Std. Err.	**Marginal Effect**	Std. Err.
ATTITUDE VARIABES				
Privacy				
Privacy is important to me	-0.018	0.019	-0.019	0.021
I would be put at a disadvantage if other could access info about my farm	-0.027*	0.016	-0.031*	0.018
I feel comfortable sharing information about my farm	0.057***	0.021	0.075***	0.020
Technology use				
I like to have the latest technology	0.008	0.016	0.014	0.021
I find new technologies easy to use	0.008	0.015	0.002	0.019
New technology is more hassle than it is worth	-0.026	0.018	-0.031*	0.018
I am getting maximum use out of available tech on my farm	-0.002	0.020	-0.007	0.018
Farm management				
I have implemented new techniques that have been recommended	0.029	0.022	0.026	0.024
I am proactive in seeking advice	0.019	0.020	0.018	0.024
Precision ag will transform agriculture over the next 20 years	0.012	0.015	0.019	0.017
I know better than others how to manage risk on my farm	-0.030*	0.017	-0.030*	0.018
REVENUE RANGE (base = $100,000-$499,999)				
<$100,000	0.027	0.117	0.069	0.146
$500,000 to $999,999	0.070*	0.041	0.097**	0.054
$1 million to $2 million	0.040	0.049	0.045	0.059
$2 million to $3 million	-0.011	0.076	0.029	0.099
>$3 million	-0.089	0.074	-0.092	0.082
COMPENSATION FOR SURVEY COMPLETION (base = $10)				
$20	-0.009	0.032	-0.013	0.040
$30	-0.015	0.038	-0.026	0.046

INTERACTION TERMS

Financial Compensation	**Revenue Range (base = $100,000-$499,999)**				
-$50	<$100,000	-0.186	0.240	-0.145	0.133
	$500,000 to $999,999	0.018	0.041	0.013	0.051
	$1 million to $2 million	0.053	0.049	0.069	0.057
	$2 million to $3 million	0.207	0.198	0.119	0.076
	>$3 million	0.024	0.116	-0.011	0.127
$50	<$100,000	-0.005	0.161	-0.057	0.162
	$500,000 to $999,999	-0.052	0.039	-0.088**	0.048
	$1 million to $2 million	0.006	0.068	-0.031	0.052

		Heteroscedastic Probit		Random Effects Probit	
Log pseudolikelihood		*-2988.39*		*-2532.71*	
Number of observations		*5265*		*5265*	
		Marginal Effect	Std. Err.	**Marginal Effect**	Std. Err.
	$2 million to $3 million	-0.016	0.213	-0.108	0.085
	>$3 million	-0.106	0.099	-0.153	0.122
	<$100,000	-0.026	0.154	-0.061	0.133
	$500,000 to $999,999	-0.007	0.054	-0.033	0.053
$100	$1 million to $2 million	-0.006	0.072	-0.033	0.055
	$2 million to $3 million	0.179	0.306	-0.054	0.094
	>$3 million	-0.176**	0.084	-0.198**	0.106

Significance codes: '' 10%, '**' 5%, '***' 1%*

Table A1.4: Variables and coefficients impacting variance for heteroscedastic probit with interaction terms between revenue and financial incentive

	Coef.	Std. Err
CHOICE VARIABLES		
Organization (base = government)		
University Researchers	-0.412*	0.249
Crop Input Suppliers	-0.438*	0.235
Equipment Manufacturers	-0.263	0.2
Grower Associations	-0.476**	0.211
Financial Institutions	-0.191	0.229
Non-Financial Incentive (base = none)		
Benchmarks	0.004	0.094
Prescription Maps	0.250**	0.103
Financial Incentive (base = $0)		
-$50	0.077	0.133
$50	0.034	0.149
$100	0.207	0.153
ATTITUDE STATEMENTS		
Privacy is important to me	0.015	0.177
I would be put at a disadvantage if other could access info about my farm	-0.085	0.086
I feel comfortable sharing information about my farm	0.019	0.154
I know better than others how to manage risk on my farm	-0.091	0.151
REVENUE RANGE (base = $100,000 to $499,999)		
<$100,000	0.472	0.808
$500,000 to $999,999	-0.017	0.248
$1 million to $2 million	0.129	0.35
$2 million to $3 million	1.26*	0.698
>$3 million	-0.144	0.493
COMPENSATION FOR SURVEY COMPLETION (base = $10)		
$20	0.197	0.238
$30	0.469	0.34

Table A1.5: Specification 3: Factor Analysis

	Heteroscedastic Probit		Random Effects Probit	
Log pseudolikelihood	*-3016.73*		*-2550.57*	
Number of observations	*5265*		*5265*	
CHOICE VARIABLES	**Marginal Effect**	Std. Err.	**Marginal Effect**	Std. Err.
Organization (base = government)				
University Researchers	0.262***	0.024	0.343***	0.024
Crop Input Suppliers	0.175***	0.022	0.221***	0.024
Equipment Manufacturers	0.074***	0.018	0.083***	0.021
Grower Associations	0.156***	0.021	0.197***	0.023
Financial Institutions	0.052***	0.019	0.065***	0.022
Non-Financial Incentive (base = none)				
Benchmarks	0.111***	0.014	0.152***	0.016
Prescription Maps	0.078***	0.014	0.108***	0.015
Financial Incentive (base = $0)				
-$50	-0.017	0.016	-0.018	0.018
$50	0.110***	0.019	0.155***	0.020
$100	0.162***	0.019	0.221***	0.022
TECHNOLOGY USE VARIABLES (base= I do not use this technology)				
Yield Monitors				
I use this technology and it improves my farm's performance	0.014	0.037	-0.002	0.050
I use this technology but it doesn't improve my farm's performance	-0.029	0.035	-0.041	0.047
GPS Guidance				
I use this technology and it improves my farm's performance	0.002	0.072	0.034	0.089
I use this technology but it doesn't improve my farm's performance	-0.056	0.089	-0.066	0.114
Soil Sampling				
I use this technology and it improves my farm's performance	0.053	0.032	0.088**	0.040
I use this technology but it doesn't improve my farm's performance	0.054	0.052	0.07	0.074
Variable rate technology				
I use this technology and it improves my farm's performance	0.035	0.036	0.022	0.042
I use this technology but it doesn't improve my farm's performance	-0.071	0.049	-0.041	0.070
Automatic Section Control				
I use this technology and it improves my farm's performance	-0.089***	0.029	-0.083**	0.039
I use this technology but it doesn't improve my farm's performance	0.283***	0.096	0.364***	0.123

	Heteroscedastic Probit		Random Effects Probit	
Log pseudolikelihood	*-3016.73*		*-2550.57*	
Number of observations	*5265*		*5265*	
	Marginal Effect	Std. Err.	**Marginal Effect**	Std. Err.
ATTITUDE VARIABES				
Privacy (increases with increasing strictness in privacy preferences)	-0.055***	0.010	-0.076***	0.012
Technology use (increases with more favourable attitudes towards technology)	0.021	0.014	0.021	0.017
Farm management (increases with increasing progressiveness in farm management style)	0.025*	0.013	0.025	0.016
REVENUE RANGE (base = $100,000 to $499,999)				
<$100,000	0.002	0.074	0.000	0.104
$500,000 to $999,999	0.072**	0.033	0.079*	0.045
$1 million to $2 million	0.069*	0.038	0.052	0.051
$2 million to $3 million	0.043	0.062	0.018	0.087
>$3 million	-0.135***	0.048	-0.165***	0.053
COMPENSATION FOR SURVEY COMPLETION (base = $10)				
$20	-0.011	0.030	-0.015	0.040
$30	-0.014	0.038	-0.016	0.047

Significance codes: '' 10%, '**' 5%, '***' 1%*

Table A1.6: Variables and coefficients impacting variance for heteroscedastic probit with factor analysis for attitude variables

	Coef.	Std. Err
CHOICE VARIABLES		
Organization (base = government)		
University Researchers	-0.278	0.266
Crop Input Suppliers	-0.372	0.228
Equipment Manufacturers	-0.245	0.19
Grower Associations	-0.379*	0.213
Financial Institutions	-0.119	0.228
Non-Financial Incentive (base = none)		
Benchmarks	0.026	0.1
Prescription Maps	0.251**	0.107
Financial Incentive (base = $0)		
-$50	0.057	0.123
$50	0.056	0.142
$100	0.184	0.139
ATTITUDE STATEMENTS		
Privacy is important to me	0.065	0.074
I would be put at a disadvantage if other could access info about my farm	-0.049	0.07
I feel comfortable sharing information about my farm	0.091	0.087
I know better than others how to manage risk on my farm	-0.217**	0.099
REVENUE RANGE (base = $100,000 to $499,999)		
<$100,000	0.343	0.518
$500,000 to $999,999	-0.049	0.18
$1 million to $2 million	0.118	0.206
$2 million to $3 million	1.25*	0.675
>$3 million	-0.132	0.371
COMPENSATION FOR SURVEY COMPLETION (base = $10)		
$20	0.193	0.164
$30	0.496	0.307

Table A1.7: Marginal Effects for Quartile Analysis; Heteroscedastic Probit

	Quartile 1		Diff. between Q1 and Q4	
	Marginal Effect	Std. Err.	**Marginal Effect**	Std. Err.
CHOICE VARIABLES				
Organization (base = government)				
University Researchers	0.302***	0.043	-0.058	0.065
Crop Input Suppliers	0.203***	0.045	-0.050	0.074
Equipment Manufacturers	0.030	0.040	0.004	0.068
Grower Associations	0.199***	0.041	-0.063	0.062
Financial Institutions	0.052	0.038	0.018	0.071
Non-Financial Incentive (base = none)				
Benchmarks	0.085***	0.030	0.015	0.039
Prescription Maps	0.046	0.031	0.051	0.049
Financial Incentive (base = $0)				
-$50	-0.046	0.032	0.090	0.056
$50	0.098**	0.044	0.068	0.058
$100	0.202***	0.047	-0.046	0.070
ATTITUDE VARIABES				
Privacy				
Privacy is important to me	-0.002	0.023	-0.052	0.043
I would be put at a disadvantage if other could access info about my farm	-0.062**	0.032	0.098**	0.039
I feel comfortable sharing information about my farm	0.053*	0.031	0.034	0.044
Technology use				
I like to have the latest technology	0.016	0.033	-0.006	0.049
I find new technologies easy to use	0.059**	0.030	-0.071*	0.041
New technology is more hassle than it is worth	-0.010	0.025	0.002	0.036
I am getting maximum use out of available tech on my farm	0.011	0.029	-0.029	0.039
Farm management				
I have implemented new techniques that have been recommended	0.024	0.037	-0.072	0.050
I am proactive in seeking advice	-0.012	0.035	0.038	0.050
Precision ag will transform agriculture over the next 20 years	0.007	0.029	0.008	0.038
I know better than others how to manage risk on my farm	-0.062**	0.031	0.050	0.048

Significance codes: '' 10%, '**' 5%, '***' 1%*
Number of observations: 2,643
Log-pseudolikelihood: -1513.77

Table A1.8: Marginal Effects for Quartile Analysis; Random Effects Probit

	Quartile 1		Diff. between Q1 and Q4	
	Marginal Effect	Std. Err.	**Marginal Effect**	Std. Err.
CHOICE VARIABLES				
Organization (base = government)				
University Researchers	0.335***	0.047	-0.035	0.065
Crop Input Suppliers	0.229***	0.048	-0.032	0.068
Equipment Manufacturers	0.034	0.044	0.019	0.078
Grower Associations	0.206***	0.045	-0.034	0.064
Financial Institutions	0.060	0.040	0.007	0.072
Non-Financial Incentive (base = none)				
Benchmarks	0.121***	0.030	0.008	0.044
Prescription Maps	0.078***	0.029	0.05	0.045
Financial Incentive (base = $0)				
-$50	-0.039	0.033	0.064	0.057
$50	0.116***	0.042	0.084	0.059
$100	0.237***	0.047	-0.036	0.057
ATTITUDE VARIABES				
Privacy				
Privacy is important to me	-0.005	0.027	-0.068	0.049
I would be put at a disadvantage if other could access info about my farm	-0.082**	0.036	0.126***	0.048
I feel comfortable sharing information about my farm	0.053	0.037	0.059	0.049
Technology use				
I like to have the latest technology	0.025	0.034	-0.006	0.057
I find new technologies easy to use	0.058	0.037	-0.08	0.051
New technology is more hassle than it is worth	-0.025	0.032	0.007	0.047
I am getting maximum use out of available tech on my farm	0.008	0.034	-0.042	0.044
Farm management				
I have implemented new techniques that have been recommended	0.039	0.044	-0.103*	0.059
I am proactive in seeking advice	0.007	0.041	0.042	0.066
Precision ag will transform agriculture over the next 20 years	0.006	0.035	0.013	0.047
I know better than others how to manage risk on my farm	-0.073**	0.037	0.055	0.055

Significance codes: '' 10%, '**' 5%, '***' 1%*
Number of observations: 2,643
Log-pseudolikelihood: -1276.68

Table A1.9: Latent class logit coefficients, no variables in the membership equation (Z_{1i})

	Class1		Class 2		Class 3	
Class probabilities	*32%*		*47%*		*21%*	
	Coef.	Std. Err.	**Coef.**	Std. Err.	**Coef.**	Std. Err.
Intercept	-6.65***	1.24	-2.37***	0.189	-1.04***	0.286
Organization (base = government)						
University Researchers	3.15***	1.19	2.04***	0.180	1.51***	0.330
Crop Input Suppliers	3.03**	1.19	1.18***	0.178	1.61***	0.332
Equipment Manufacturers	-25.2	353856	0.487**	0.190	1.08***	0.317
Grower Associations	2.60**	1.22	1.15***	0.178	1.28***	0.327
Financial Institutions	1.76	1.29	0.339*	0.188	0.752**	0.300
Non-Financial Incentive (base = none)						
Benchmarks	0.674	0.419	1.00***	0.121	0.778***	0.228
Prescription Maps	1.14***	0.400	0.692***	0.121	0.503**	0.217
Financial Incentive (base = $0)						
-$50	0.039	0.509	0.022	0.146	0.288	0.222
$50	0.853*	0.461	0.628***	0.144	2.64***	0.518
$100	1.03**	0.452	1.18***	0.146	1.97***	0.318

Significance codes: '' 10%, '**' 5%, '***' 1%*
Number of observations: 5265
Log-likelihood: -2588.20

Table A1.10: Latent class overall partial effects, no variables in the membership equation (Z_{1i})

	Partial Effect	Elasticity
Organization (base = government)		
University Researchers	0.167***	0.351
Crop Input Suppliers	0.136***	0.286
Equipment Manufacturers	-0.557	-1.17
Grower Associations	0.120***	0.253
Financial Institutions	0.064**	0.135
Non-Financial Incentive (base = none)		
Benchmarks	0.062***	0.262
Prescription Maps	0.058***	0.244
Financial Incentive (base = $0)		
-$50	0.006	0.019
$50	0.082***	0.257
$100	0.094***	0.299

Significance codes: '' 10%, '**' 5%, '***' 1%*
Number of observations: 5265

Table A1.11: Latent class logit coefficients, attitude statements in the membership equation (Z_{3i})

	Class1		Class 2		Class 3	
Class probabilities	*32%*		*19%*		*49%*	
	Coef.	Std. Err.	**Coef.**	Std. Err.	**Coef.**	Std. Err.
Intercept	-6.76***	1.223	-0.954***	0.319	-2.27***	0.199
Organization (base = government)						
University Researchers	3.39***	1.15	1.46***	0.352	2.00***	0.183
Crop Input Suppliers	3.01**	1.18	1.53***	0.352	1.22***	0.176
Equipment Manufacturers	1.23	1.29	1.04***	0.344	0.476**	0.189
Grower Associations	2.43**	1.19	1.22***	0.340	1.21***	0.175
Financial Institutions	2.13*	1.21	0.743**	0.315	0.302	0.187
Non-Financial Incentive (base = none)						
Benchmarks	0.873**	0.390	0.803***	0.242	0.980***	0.122
Prescription Maps	1.19***	0.371	0.474**	0.227	0.672***	0.120
Financial Incentive (base = $0)						
-$50	0.170	0.436	0.341	0.246	-0.003	0.146
$50	0.916**	0.397	2.92***	0.641	0.637***	0.149
$100	1.11***	0.401	2.01***	0.331	1.19***	0.149

Significance codes: '' 10%, '**' 5%, '***' 1%*

Number of observations: 5265

Log-likelihood: -2552.04

Table A1.12: Coefficients impacting class membership, attitude statements in the membership equation (Z_{3i}), base = class 3

	Class 1		Class 2	
	Coef.	Std. Err.	**Coef.**	Std. Err.
Intercept	0.208	1.09	-1.10	1.28
Privacy				
Privacy is important to me	0.020	0.153	-0.123	0.167
I would be put at a disadvantage if other could access info about my farm	0.147	0.143	-0.139	0.157
I feel comfortable sharing information about my farm	-0.595***	0.166	0.046	0.198
Technology use				
I like to have the latest technology	-0.031	0.174	-0.085	0.199
I find new technologies easy to use	-0.164	0.156	-0.192	0.180
New technology is more hassle than it is worth	0.288**	0.135	-0.072	0.160
I am getting maximum use out of available tech on my farm	0.425***	0.147	0.432**	0.175
Farm management				
I have implemented new techniques that have been recommended	-0.280	0.185	-0.054	0.215
I am proactive in seeking advice	-0.114	0.181	-0.045	0.216
Precision ag will transform agriculture over the next 20 years	-0.010	20.7	0.243	27.9
I know better than others how to manage risk on my farm	0.229	63.4	0.003	76.6

Significance codes: '' 10%, '**' 5%, '***' 1%*

Number of observations: 5265

Table A1.13: Latent class overall partial effects, attitude statements in the membership equation (Z_{3i})

	Partial Effect	Elasticity
Organization (base = government)		
University Researchers	0.464***	0.285
Crop Input Suppliers	0.367***	0.225
Equipment Manufacturers	0.163**	0.100
Grower Associations	0.317***	0.195
Financial Institutions	0.193**	0.118
Non-Financial Incentive (base = none)		
Benchmarks	0.181***	0.222
Prescription Maps	0.158***	0.194
Financial Incentive (base = $0)		
-$50	0.023	0.021
$50	0.228***	0.210
$100	0.260***	0.240

Significance codes: '' 10%, '**' 5%, '***' 1%*

Number of observations: 5265

Table A1.14: Latent class logit coefficients, technology use variables in the membership equation (Z_{4i})

	Class1		Class 2		Class 3	
Class probabilities	*29%*		*19%*		*52%*	
	Coef.	Std. Err.	**Coef.**	Std. Err.	**Coef.**	Std. Err.
Intercept	-6.45***	1.26	-1.00***	0.339	-2.26***	0.196
Organization (base = government)						
University Researchers	3.18***	1.18	1.44***	0.381	2.03***	0.181
Crop Input Suppliers	2.89**	1.21	1.60***	0.376	1.21***	0.179
Equipment Manufacturers	-0.074	3.01	0.898***	0.340	0.570***	0.183
Grower Associations	2.52**	1.22	1.35***	0.365	1.15***	0.177
Financial Institutions	1.69	1.25	0.627**	0.319	0.401**	0.181
Non-Financial Incentive (base = none)						
Benchmarks	0.716*	0.387	0.842***	0.256	0.971***	0.120
Prescription Maps	1.09***	0.352	0.514**	0.230	0.665***	0.120
Financial Incentive (base = $0)						
-$50	0.142	0.436	0.501**	0.252	-0.069	0.143
$50	0.873**	0.400	2.93***	0.687	0.628***	0.157
$100	1.02**	0.415	2.06***	0.333	1.17***	0.151

Significance codes: '' 10%, '**' 5%, '***' 1%*

Number of observations: 5265

Log-likelihood: -2568.62

Table A1.15: Coefficients impacting class membership, technology use variables in the membership equation (Z_{4i}), base = class 3

	Class 1		Class 2	
	Coef.	Std. Err.	**Coef.**	Std. Err.
Intercept	-0.178	0.680	-1.90*	1.106
Yield Monitors				
I use this technology and it improves my farm's performance	-0.335	0.369	0.075	0.522
I use this technology but it doesn't improve my farm's performance	0.532	0.353	0.779*	0.472
GPS Guidance				
I use this technology and it improves my farm's performance	-0.071	0.637	0.556	0.994
I use this technology but it doesn't improve my farm's performance	1.150	0.977	0.914	1.278
Soil Sampling				
I use this technology and it improves my farm's performance	-0.334	0.326	0.635	0.487
I use this technology but it doesn't improve my farm's performance	0.017	0.492	0.608	0.689
Variable rate technology				
I use this technology and it improves my farm's performance	0.257	0.333	0.640	0.433
I use this technology but it doesn't improve my farm's performance	0.911*	0.524	0.040	0.751
Automatic Section Control				
I use this technology and it improves my farm's performance	-0.078	0.287	-1.04***	0.355
I use this technology but it doesn't improve my farm's performance	-29.9***	9.72	-0.167	12.2

Significance codes: '' 10%, '**' 5%, '***' 1%*

Number of observations: 5265

Table A1.16: Latent class overall partial effects, technology use in the membership equation (Z_{4i})

	Partial Effect	Elasticity
Organization (base = government)		
University Researchers	0.461***	0.267
Crop Input Suppliers	0.362***	0.210
Equipment Manufacturers	0.091	0.053
Grower Associations	0.324***	0.189
Financial Institutions	0.167**	0.097
Non-Financial Incentive (base = none)		
Benchmarks	0.179***	0.208
Prescription Maps	0.155***	0.180
Financial Incentive (base = $0)		
-$50	0.021	0.018
$50	0.232***	0.203
$100	0.265***	0.231

Significance codes: '' 10%, '**' 5%, '***' 1%*

Number of observations: 5265

APPENDIX 2: Model Choice and Construction

Heteroscedastic Probit

In a heteroscedastic probit model, the variance is free to vary with a set of variables. This is contrary to a probit model where variance is fixed at one. The variables that variance moves with are contained in Z, and may or may not bee a subset of X. The probability of success (the probability that the farmer will participate in a big data program) for the *i*th observation is:

$$\Pr(y_i = 1|x_i, z_i) = \Phi\left\{\frac{x_i b}{\exp(z_i \gamma)}\right\},$$

where $y_i = 1$ if the respondent chooses to participate in the big data program,
x_i is a vector of variables that impact y_i,
z_i is a vector of variables that impact the variance, and
b, γ are coefficient vectors.

The probability of failure is $\Pr(y_i = 0|x_i, z_i) = 1 - \Pr(y_i = 1|x_i, z_i)$. Given this, the density for each observation is:

$$f(y_i|x_i, z_i) = \Pr(y_i = 1|x_i, z_i)^{y_i} * \Pr(y_i = 0|x_i, z_i)^{1-y_i}$$

$$f(y_i|x_i, z_i) = \left[\Phi\left\{\frac{x_i b}{\exp(z_i \gamma)}\right\}\right]^{y_i} * \left[1 - \Phi\left\{\frac{x_i b}{\exp(z_i \gamma)}\right\}\right]^{1-y_i}$$

Multiplying the density for each function and taking the natural logarithm gives the log-likelihood function:

$$L(\beta, \Gamma) = \sum_{i=1}^{n}\left\{y_i * ln\left[\Phi\left\{\frac{x_i \beta}{\exp(z_i \Gamma)}\right\}\right] + (1 - y_i) * ln\left[1 - \Phi\left\{\frac{x_i \beta}{\exp(z_i \Gamma)}\right\}\right]\right\}$$

Optimizing the log-likelihood with respect to β and Γ gives their maximum likelihood estimates, b and γ. Standard errors are clustered on respondent (each respondent accounts for approximately 12 observations).

To determine the variables that influence variance (Z), we performed a Breusch-Pagan test. We obtained the squared residuals from a probit model with all explanatory variables (X) included and regressed them on the explanatory variables (X). In the resulting output, the variables that emerged as statistically significant at the 10% level were determined to impact the variance. These variables were chosen to make up Z. These variables were organization, financial incentive, non-financial incentive, 'Privacy is important to me', 'I would be put at a disadvantage if others could access info about my farm', 'I feel comfortable sharing information about my farm', 'I know better than others how to manage risk on my farm', farm revenue, and the incentive provided to respondents for completing the survey.

Random Effects Probit

The second model used is a random effects probit model. A traditional random effects model uses generalized least squares (GLS) to generate estimates, however, the random effects probit model uses maximum likelihood. This is because maximum likelihood is required for the 'probit' part of the analysis. Observations are grouped by respondent (or by panel). The panel-level likelihood function is,

$$l_i = \int_{-\infty}^{\infty} \frac{e^{-v_i^2/2\sigma_v^2}}{\sqrt{2\pi}\sigma_v} \left\{ \prod_{q=1}^{n_i} F(y_{iq}, x_{iq}\beta + v_i) \right\} dv_i$$

where:

$$F(y_{iq}, x_{iq}\beta + v_i) = \begin{cases} \Phi(x_{iq}\beta + v_i) \quad if \quad y = 1 \\ 1 - \Phi(x_{iq}\beta + v_i) \quad if \quad y = 0 \end{cases}$$

and: v_i is the random effect,
i is an index for individual,
q is an index for question number,
n_i is the number of responses for the *i*th individual, and
Φ is the cumulative normal distribution.

The log-likelihood function is the sum of the logs of the panel-level likelihoods. Maximizing the log-likelihood function gives the estimates for β and σ_v. The random effects probit model analyzes variation between individuals (between variation) and variation between observations from the same individual (within variation).

Factor Analysis

The attitude variables are analyzed using factor analysis. Factor analysis can be useful where there is correlation between a set of variables included in a model. Factor analysis attempts to condense the overall variation in a set of variables into one measure. It is appropriate for the attitude statements because there are multiple statements attempting to capture the same measure.

The first three statements are attempting to capture privacy attitudes, the next four are capturing technology use attitudes, and the final four are capturing farm management attitudes. Because they are capturing the same thing, high correlation between variables in the same category is expected. The correlation matrix for the eleven attitude variables is shown in table A2.1.. High numbers are expected within the boxes and low numbers are expected outside the boxes. This is generally true, however there are some high correlations between statements measure technology use attitudes and farm management attitudes, suggesting these measures may be confounded.

Table A2.1: Correlation matrix for attitude variables

		Privacy			Technology Use				Farm Management			
		1	2	3	4	5	6	7	8	9	10	11
Privacy	1	1.00										
	2	0.36	1.00									
	3	-0.24	-0.37	1.00								
Technology Use	4	0.11	0.13	0.07	1.00							
	5	0.15	0.07	0.11	0.34	1.00						
	6	0.03	0.10	-0.05	-0.27	-0.12	1.00					
	7	0.12	0.07	0.09	0.34	0.30	-0.08	1.00				
Farm Management	8	0.14	0.03	0.17	0.48	0.30	-0.13	0.28	1.00			
	9	0.17	0.08	0.13	0.39	0.40	-0.16	0.28	0.38	1.00		
	10	0.11	0.01	0.12	0.27	0.17	-0.20	0.13	0.19	0.27	1.00	
	11	0.21	0.10	0.05	0.13	0.16	0.12	0.20	0.13	0.16	0.11	1.00

Factor analysis takes the total variation from a set of variables that are correlated and breaks it up into an equal number of orthogonal factors that are used in place of the variables in the analysis. The first factor captures the most possible total variation in the variables examined. The second factor captures the most possible total variation that remains unexplained by the first factor. This process continues until the total variation in the variables examined is captured by the factors. As successive factors become less important, they can be excluded from the analysis. I perform factor analysis on each group of attitude variables. The goal of this analysis is to synthesize the responses from the statements in each category of the attitude variables, so the final model has one measure of privacy attitudes, one measure of attitudes towards technology, and one measure of farm management attitudes.

The decision rule on the number of factors to include in the analysis depends on the percentage of variation each factor captures. Any factor that captures more than its fair share of variation should be included. The fair share of variation for each factor is the amount of variation it would capture if the total variation was equally divided among all factors. The proportion of variation explained by each factor is shown in table A2.2. Only one factor from each group captured more than its fair share of variation, so only one factor for each group is included in the models. There is one measure of privacy attitudes, one measure of technology use attitudes, and one measure of farm management attitudes.

Table A2.2: Proportion of variation explained by generated factors

Privacy *(fair share = 33.3%)*		**Technology Use** *(fair share = 25%)*		**Farm Management** *(fair share = 25%)*	
Factor	Percent variation explained	Factor	Percent variation explained	Factor	Percent variation explained
1	55%	1	44%	1	41%
2	25%	2	24%	2	23%
3	20%	3	18%	3	21%
		4	15%	4	15%

Each factor has a related eigenvector where each value in the vector corresponds to a variable analyzed. This eigenvector describes the composition of the factor. Table A2.3 shows the eigenvectors associated with the first factor for privacy attitudes, attitudes towards technology use, and farm management attitudes. From these values we can determine the relationship between the variables and the factors. The signs of the values in the eigenvectors line up with expectations except for the last statement in farm management attitudes.

The factor describing privacy preferences increases with increasing strictness in privacy preferences. The factor describing attitudes towards technology increases with more favourable attitudes towards technology. The factor describing farm management attitudes increases with increasing progressiveness in farm management style. The results from factor analysis can be difficult to interpret, however broad interpretations regarding effect direction can be made.

Table A2.3: Eigenvectors for first factor for each group

Privacy	**Expected Sign**	**Eigenvector**
Privacy is important to me	+	0.55
I would be put at a disadvantage if other could access info about my farm	+	0.62
I feel comfortable sharing information about my farm	-	-0.56
Technology Use		
I like to have the latest technology	+	0.59
I find new technologies easy to use	+	0.51
New technology is more hassle than it is worth	-	-0.34
I am getting maximum use out of available tech on my farm	+	0.51
Farm Management		
I have implemented new techniques that have been recommended	+	0.55
I am proactive in seeking advice	+	0.60
Precision ag will transform agriculture over the next 20 years	+	0.46
I know better than others how to manage risk on my farm	-	0.34

Latent Class Model Selection

Two latent class models are constructed. One includes two latent classes and on includes three latent classes. To determine which model fits the data better, Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC) are used. They are calculated as,

$$AIC = -2LL + 2m \qquad \text{and} \qquad BIC = -2LL + m * \ln(n),$$

where:

LL is the log-likelihood,
m is the number of parameters calculated, and
n is the number of observations.

The last term in each expression is the penalty component serving to discourage overfitting of a model. The model with the lower AIC or BIC is preferred. Both measures conclude that the optimal number of classes for the latent class analysis is two.