Farmers' Willingness to Share Data: A Study of Saskatchewan Farmers

M. Turland[1]; P. Slade[2]

1: University of Saskatchewan, Department of Agricultural and Resource Economics, Canada,
2: University of Saskatchewan, Department of Agriculture and Resource Economics, Canada

Corresponding author email: madeline.turland@usask.ca

Abstract:

Modern farm machinery captures geocoded data on all aspects of a farming operation. These detailed datasets are called big data. 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 trends. To analyze farmers' willingness to share data we conduct a hypothetical choice experiment that asked farmers in Saskatchewan whether they would join a big data 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 data from a survey constructed for this study. 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.

Acknowledegment:

JEL Codes: Q16, C51

#1530

Farmers’ Willingness to Share Data: A Study of Saskatchewan Farmers

Modern farm machinery captures geocoded data on all aspects of a farming operation. These detailed datasets are called big data. 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 trends.

To analyze farmers’ willingness to share data we conduct a hypothetical choice experiment that asked farmers in Saskatchewan whether they would join a big data 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 data from a survey constructed for this study. 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.

INTRODUCTION

Modern agricultural technology generates detailed farm level datasets that measure exact inputs and outputs. This data is not only useful to farmers, but has the potential to transform technological innovation and production in agriculture. Having detailed datasets that measure exact farm inputs and outputs could help focus where research investment should be made, and what expenditure dollars should be spent on to see maximum improvements in yield, environmental stewardship, and automation. The detailed nature of this type of data puts it in a category labelled “big data”. To realize the benefits of big data, researchers must have access to a large enough sample to reveal underlying trends. In Canada, the size of the agricultural industry dictates that a large proportion of farmers must participate in a big data program for it to be valuable. Achieving high participation rates is a primary goal when constructing a big data program. Farmers reluctance to share information about themselves works contrary to this goal. Farmers tend to value privacy, and are wary of sharing information about themselves and their operations. This paper estimates farmers’ willingness to share big data, who they are willing to share it with, and what incentivizes them to do so.

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. Olsen, Grundin, and Horvitz (2005) attempt to classify people into groups according to their privacy preferences, and determine that privacy preferences are fixed for individuals. Lovejoy, Horn, and Hughes (2009) also find that privacy preferences are fixed. Educating people on privacy policies and potential risk (or the absence of risk) doesn't change their behaviour. If privacy preferences are fixed, then farmers will not be induced to share their data through social coercion methods such as advertising campaigns. Other methods must be used to achieve high participation rates among farmers for a big data program.

Athey, Catalini, and Tucker (2017) surveyed students at the Massachusetts Institute of Technology (MIT) and found that students are willing to relinquish data quite readily when incentivized to do so, even though their stated privacy preference may be strong. This is an example of the privacy paradox, and highlights the importance of experiments in privacy research. Relative to their actions, people tend to overstate their privacy preferences when directly asked.

The purpose of this paper is to determine the key factors that influence farmers' willingness to participate in a big data program in Saskatchewan. A heteroscedastic probit model and a random effects probit model are constructed, using data from a hypothetical choice experiment designed for this study. Respondents were asked a series of twelve choice questions. Each question asked respondents whether they would participate in a big data program, under varying conditions. The questions differed in organization that operated the program, financial reward for participating in the program, and non-financial compensation for enrollment. One characteristic from each category was randomly assigned to each question. We attempt to control for farm revenue and privacy attitudes in the analysis.

The results are consistent across models. The institution responsible for coordinating the big data program is particularly important for participation rates. Farmers are most willing to join a program run by university researchers, followed by crop input suppliers or grower organizations, then equipment manufacturers or financial institutions. Farmers are least willing to participate in a program run by government. The presence of a financial incentive has a larger impact on participation rates than the size of the financial incentive.

This paper 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 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. Agriculture is moving towards data driven innovations, and having a big database will become more valuable for improvements to productivity.

The first section below is further background surrounding issues related to big data and data management in agriculture. The following section outlines the data used for the statistical analysis. Then section four presents the empirical models used. Section five shows the results, and section six presents a discussion and conclusion.

BACKGROUND

The term "big-data" was coined by Cox and Ellsworth (1997), who defined a dataset as "big" if it was too large to be processed by traditional software. Since this time the definition of big data has become the subject of debate. George, Haas, and Pentland (2014) suggest that big data is defined by the fine-grained nature of the data itself; it doesn't 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 et al (2015) describe the three V's of big data: volume, velocity, and variety. These characteristics reveal the challenges of 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 big data analysis was prohibitively costly in most situations. With falling storage and analyzing costs, big data is becoming more popular. Private companies are seeing the benefits of investing in data, and even racing 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).

In agriculture, big data describes datasets created by sophisticated machinery that quantify inputs and outputs on a farming operation at a micro level. Big data is part of the precision agriculture revolution that aims to increase automation and productivity. Examples of big data in agriculture include yield maps generated by real time yield monitors, variable rate input application based on changing soil characteristics, and increased knowledge of how inputs affect yield and the environment.

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. Bronson and Knezevic (2016) say that the benefits of big data and precision agriculture might be oversold to farmers by the companies that make the equipment and provide services. Many precision agriculture techniques have yet to be proven, and it is uncertain whether their promised benefits will ever be realized. Advances using precision agriculture may be insignificant when it comes to increased production, as weather continues to be the most important factor. Farmers must see at least some benefits from the use of precision agriculture techniques as the use of this technology is rising. From the survey data, 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. However, it's unclear whether farmers will see any direct benefits from big data. Bronsen and Knezevic (2016) predict that most of the benefits from big data will flow to a small number of large agriculture companies.

Researchers can use big data to uncover underlying trends that were previously impossible to detect. These could be useful for increasing productivity in the agricultural sector. Researchers can identify in which areas the greatest gains can be made from investment. For example, a researcher would be able to monitor the effects of fertilizer use on environmental degradation, including changes in soil salinity. The reality is that many of the benefits that come from using big data are unknown. Researchers don't yet know what they will be able to accomplish with a big dataset. We do know that data brings information and opportunity.

Big data brings opportunities for researchers, but it also brings challenges. George, Haas, and Pentland (2014) discuss how the sheer number of observations in the dataset could reveal false correlations. Working with big datasets 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. Using big data also means accepting messiness. 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 shouldn't 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).

Pavolotsky (2013) identifies a unique problem with sharing big data. The value of big data lies in identifying secondary uses of the data that remain 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 practical limits of meaningful consent. This could be an important factor moving forward in the world of big data.

In agriculture, there are many stakeholders that see the value in constructing a big dataset. Government has vested interests in greater agricultural production and higher farm incomes.

University researchers want the data purely for academic reasons. Equipment manufacturers and crop input suppliers are continually attempting to innovate new products. 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. These organizations could all conceivably decide to pioneer a big data program as all would see benefits.

As a databank becomes larger, it not only becomes more valuable, but it also becomes more widely known among farmers, and will likely generate higher response rates. 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 may have an advantage because they can observe and learn from the first-to-market's model, and develop a better project. Whatever the case, the organization that decides to invest in the construction of a big data program will determine the structure of the database, including who has 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."

Farmers have access to incredible amounts of data. The full value of this data is not extracted by the farmer because analyzing it is costly, and the value from aggregation is lost. Researchers would like to harness the power of big data in agriculture, but cannot do so until a databank capturing big data from farms across the country exists.

DATA

The data comes from a survey on farmers' willingness to pay. The survey was administered online by Kynetec 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 the 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 any statistically significant impact on the results. Overall the survey generated 561 responses.

The question used to formulate the outcome variable was a binary response question asking if the respondent would participate in a big data program under specific conditions. Each question presented a scenario that combined one organization, one financial incentive, and one non-financial incentive. The options studied are presented in table 1. Each survey respondent was asked to evaluate twelve separate scenarios. These questions are referred to as the choice questions in this paper. Figure 1 is a screenshot of the information script proceeding the choice question set. Figure 2 is an example of a choice question presented in the survey.

Table 1: Organizations and incentives studied

Organization	Financial Incentive	Non-Financial Incentive
1. University Researchers 2. Crop Input Suppliers 3. Grower Associations 4. Equipment Manufacturers 5. Financial Institutions	1. -\$50 2. \$0 3. \$50 4. \$100	1. Prescription maps based on the data submitted. Depending on the data submitted these could be for fertilizer, seed, fungicide, or other inputs. 2. 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."

Figure 1: Information script preceding choice questions

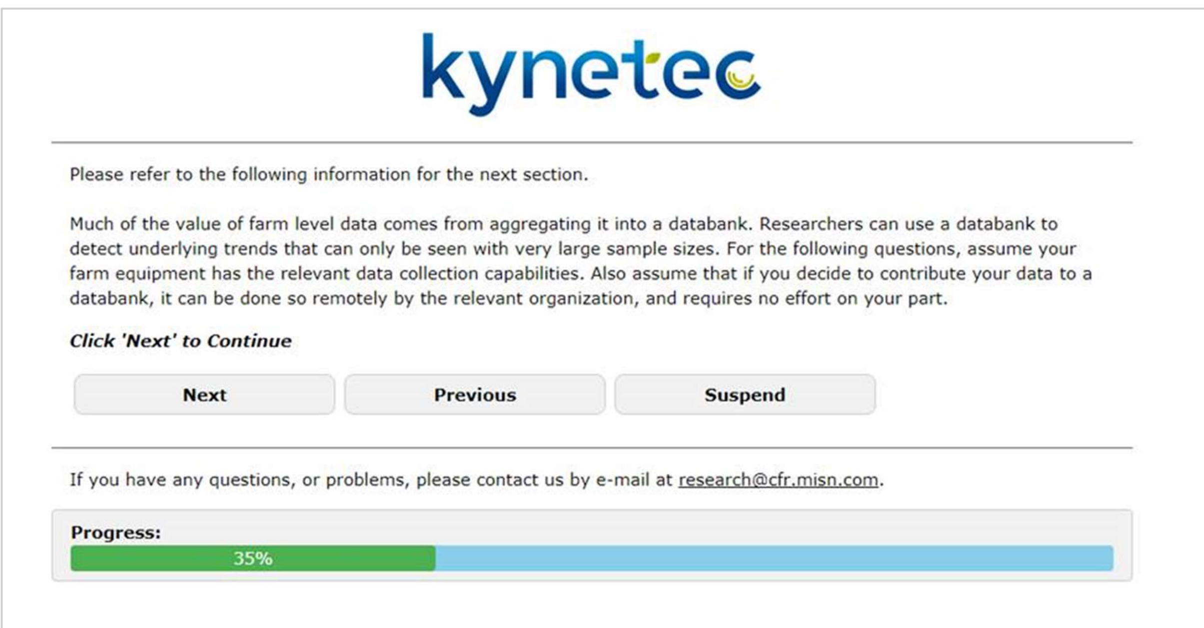
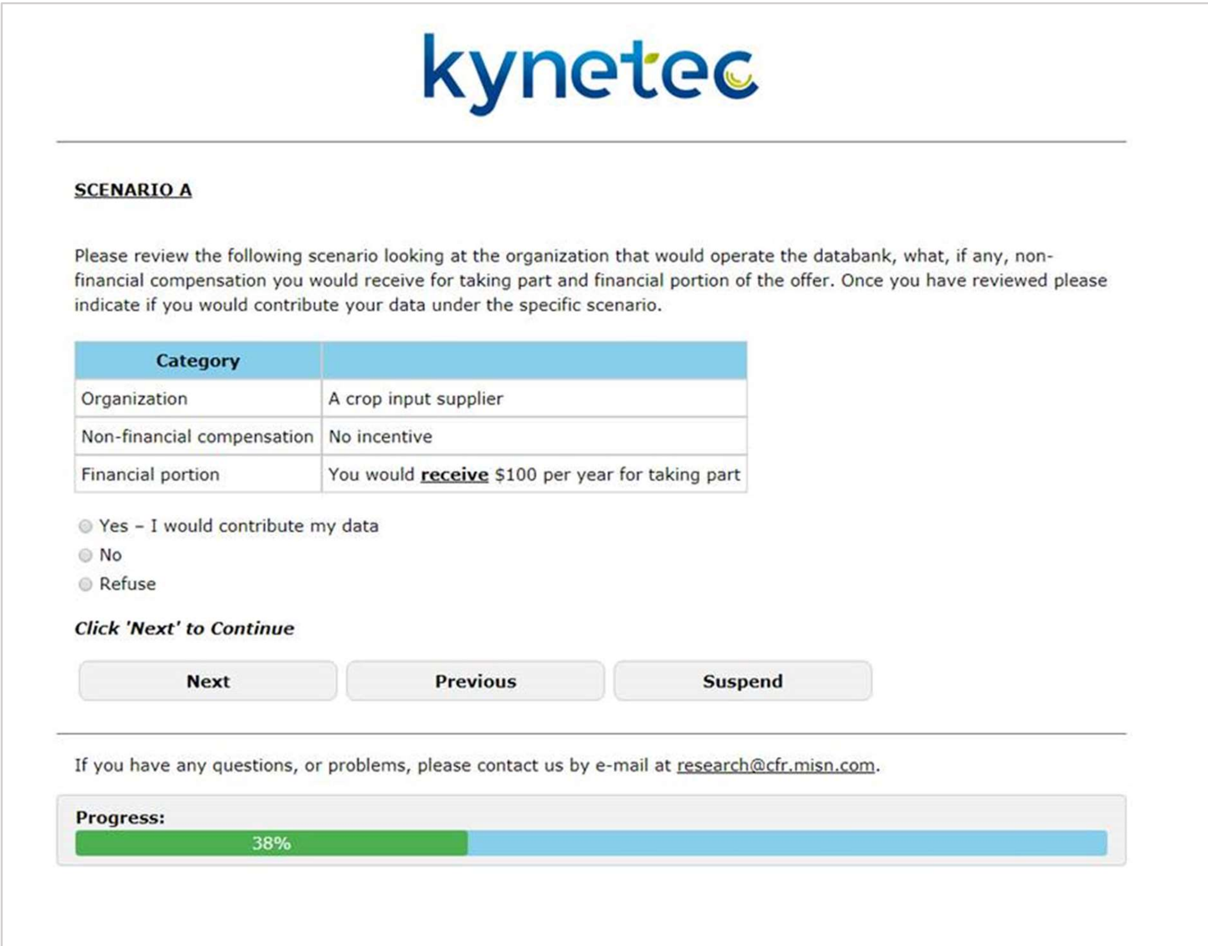


Figure 2: Choice question example


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%

The survey design was pseudo-random. Seventy-two unique scenarios are possible with six organizations, four financial incentives, and three non-financial incentives. Each respondent was asked to evaluate only twelve of these scenarios. The seventy-two unique scenarios were divided into six groups of twelve, ensuring sufficient variation 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 are subdivided into statements concerning attitudes towards privacy (first three statements), technology use (next four statements), and farm management (final four statements).

The statements, their mean responses, and standard deviations are shown in table 2. These questions are referred to as the attitude questions in this paper.

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”. These questions are referred to as the technology use questions. Additional questions captured descriptive information about the individual and operation, including demographic information, farm financial information, and farm structure information.

The survey resulted in a panel dataset. Each respondent answered twelve choice questions, which yielded twelve observations for the study. The sample size is not 561, but rather 6732. However, not all respondents answered all questions as there was a “refuse” option. Table 3 shows how the analysis sample was constructed. Observations with missing covariates are removed. This doesn’t bias the results as they are consistent when dummy variables are added for non-response. The final analysis sample includes no missing covariates, and has 5571 observations.

Overall, farmers were willing to participate in a big data program 36% of the time. Figure 3 shows how the participation rate changes for organization, financial incentive, and non-financial incentive.

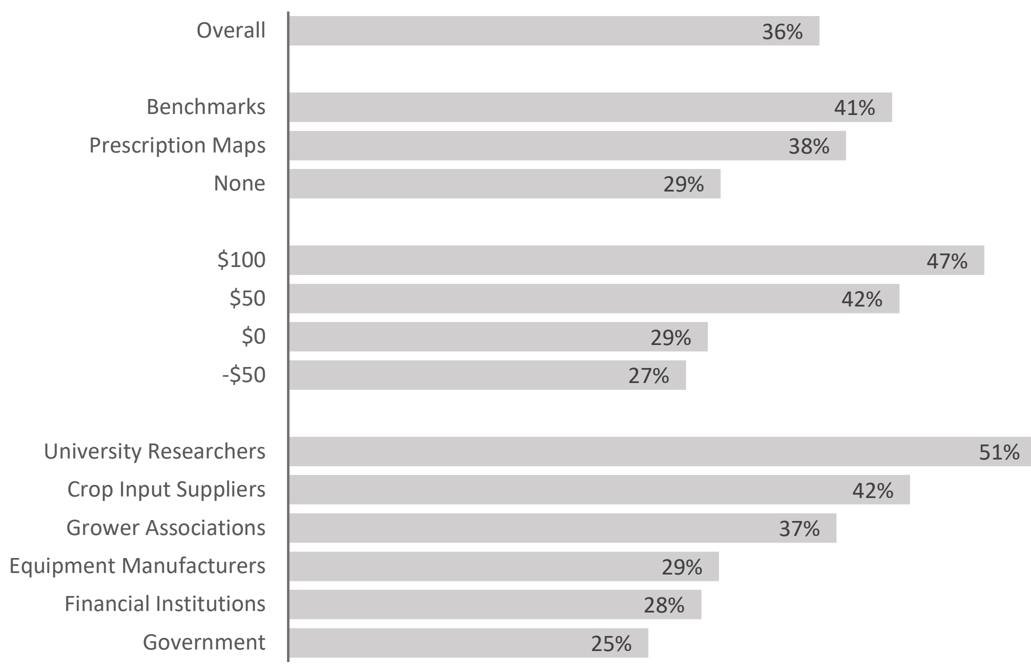
Table 2: Respondents rate their level of agreement on a scale of 1 to 5

	Mean	St. Dev.
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
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
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

Table 3: Analysis Sample Construction

Starting sample size	6732
Less missing choice covariates	6510
Less missing technology use covariates	6419
Less missing attitude covariates	6245
Less missing income covariates (analysis sample)	5571

Figure 3: Big Data Program Participation Rates for Select Subsamples



EMPIRICAL MODEL

This paper uses a heteroscedastic probit model, and a random effects probit model to analyze farmers willingness to share data. The discrete variables included in the models are for organization, financial incentive, non-financial incentive, technology use, revenue range, and compensation for completing the survey. The attitude questions are included as continuous variables. The dependent variable is constructed from the responses to the choice questions.

A Breusch-Pagan test was used to determine if heteroscedasticity was present in the data. The results indicated that the error term varied with organization, financial incentive, and non-financial incentive. In addition, there was heteroscedasticity for GPS guidance and variable rate technology under the technology use questions. For the attitude questions, “Privacy is important to me”, “I feel comfortable sharing information about my farm”, and “I know better than others how to manage risk on my farm” experienced heteroscedasticity. These variables were controlled for in the heteroscedastic probit model.

A random effects probit model was chosen to account for correlation between multiple observations from the same individual. It allows for descriptive statistics that are constant for individual across their responses (such as farm revenue) to be included in the model.

In the heteroscedastic probit model, standard errors were clustered on respondent to account for correlation between responses by the same individual. In the random effects probit model, robust standard errors were reported.

RESULTS

The marginal effects and standard errors for the heteroscedastic probit and random effects probit models are shown in table 4. A chi-square test measures if coefficients are statistically different from each other. Coefficients that are not statistically different from each other indicate that farmers are indifferent between participating in a big data program run by the relevant organizations.

The results for organization, financial incentive, and non-financial incentive are consistent across models. When compared to government, farmers are 27% (35%) more likely to share their data with university researchers in the heteroscedastic (random effects) probit model. Farmers are 17% (21%) more likely to share with crop input suppliers, 16% (19%) more likely to share with grower associations, 6% (7%) more likely to share with equipment manufacturers, and 5% (6%) more likely to share with financial institutions than with government in the heteroscedastic (random effects) model. There is no statistical difference between the size of the marginal effects for crop input suppliers and grower associations, or for equipment manufacturers and financial institutions, suggesting indifference between these organizations. This result is consistent with figure 3 above.

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% (16%) more likely to participate if offered $50, and 18% (23%) more likely to participate if offered $100. Paying $50 for the right to participate has no statistically significant effect on farmer’s willingness to participate. This is surprising considering the statistical strength of the coefficients related to positive financial compensation.

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). Even if people have strong privacy preferences, they can be induced to share information for remarkably small rewards.

The difference between the marginal effects for the non-financial incentives is statistically significant at the five percent level but not the one percent level for both models. In the heteroscedastic (random effects) probit model, farmers are 10% (14%) more likely to participate in a big data program if offered benchmark statistics in return, and 7% (11%) more likely to participate if offered prescription maps as compensation. Prescription maps are only useful to those farmers that use variable rate technology (29% of our sample), while benchmark statistics are useful to every farmer regardless of the equipment available to them. This results also suggests that farmers not only care about their individual productivity, but also are interested in how they perform in relation to other farmers.

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" could have revealed a statistically significant relationship with farmers' willingness to share data that doesn't economically exist. Forty-eight observations over four individuals is too small of a sample to draw a decisive conclusion. There is not strong evidence supporting a relationship between technology use and farmer's willingness to share data in Saskatchewan.

We analyze the attitude questions in terms of their grouping into statements on privacy (first three statements), technology use (next four statements), and farm management (final four statements). To test the overall impacts of privacy, technology use, and farm management, we performed joint significance tests for the statements in each group. The findings did not reveal any relationships in the data that were not present when individual tests for significance were performed.

Attitudes concerning privacy had the largest impact on farmers' willingness to share. The only statement 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.

None of the technology attitude statements are statistically significant in the heteroscedastic probit model. In the random effects probit model "New technology is more hassle than it is worth" had a negative statistically significant coefficient. It is intuitive that those that struggle to see the value of new technology are less willing to adopt that technology. Those that see less value in new technology also likely place less value on innovation. They are less willing to participate in a big data program because they don't believe the program will yield real world benefits. This result is not robust across models, and is weak overall.

None of the farm management statements were statistically significant in the random effects probit model, however in the heteroscedastic probit model two of the statements were. "I know better than others how to manage risk on my farm" had a negative effect, and "I am proactive in seeking advice" had a positive effect. These results suggest that as a farmer has a more independent managing style, they become less willing to participate in a big data program. Farmers that believe they can improve their operations by seeking outside council are more willing to share their data.

Farmers whose operations generate the most revenue (>$3 million/year) are less likely to share their data than those that have farm revenue under $100 000. Farmers running larger operations won't 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. Farmers generating less than $100 000 in revenue each year likely generate some off farm income to supplement their farm income. In additions, farms that generate more revenue are more likely to be incorporated. In the analysis sample, 100% of farms with sales revenue over $3 million were incorporated while 24% of farms with sales revenue under $100 000 were incorporated.

The different incentives provided to farmers to complete the study had no impact on their willingness to share data in the analysis. This is a little surprising, considering financial incentive had statistically significant impacts in the results of the analysis on farmers' willingness to share data. This suggests that small monetary rewards can influence farmers' behaviour. It is not only whether a financial incentive exists that is important, but the size of the incentive is important as well.

Table 4: Results for Heteroscedastic Probit and Random Effects Probit

	Heteroscedastic Probit		Random Effects Probit	
	Marginal Effect	**Std. Err.**	**Marginal Effect**	**Std. Err.**
CHOICE VARIABLES				
Organization (base = government)				
University Researchers	0.267***	0.021	0.349***	0.023
Crop Input Suppliers	0.170***	0.020	0.214***	0.023
Equipment Manufacturers	0.056***	0.018	0.071***	0.021
Grower Associations	0.158***	0.019	0.188***	0.022
Financial Institutions	0.045**	0.019	0.06***	0.021
Non-Financial Incentive (base = none)				
Benchmarks	0.099***	0.013	0.143***	0.016
Prescription Maps	0.070***	0.013	0.105***	0.015
Financial Incentive (base = $0)				
-$50	-0.017	0.016	-0.016	0.017
$50	0.123***	0.016	0.156***	0.019
$100	0.176***	0.018	0.229***	0.021
TECHNOLOGY USE VARIABLES (base = I do not use this technology)				
Yield Monitors				
I use this technology and it improves my farm’s performance	0.029	0.040	0.021	0.047
I use this technology but it doesn’t improve my farm’s performance	-0.014	0.037	-0.019	0.044
GPS Guidance				
I use this technology and it improves my farm’s performance	0.007	0.048	0.022	0.081
I use this technology but it doesn’t improve my farm’s performance	-0.055	0.069	-0.073	0.105
Soil Sampling				
I use this technology and it improves my farm’s performance	0.031	0.033	0.047	0.039
I use this technology but it doesn’t improve my farm’s performance	0.033	0.053	0.061	0.071
Variable rate technology				
I use this technology and it improves my farm’s performance	0.022	0.037	0.021	0.041
I use this technology but it doesn’t improve my farm’s performance	-0.058	0.056	-0.022	0.072
Automatic Section Control				
I use this technology and it improves my farm’s performance	-0.071**	0.028	-0.068*	0.037
I use this technology but it doesn’t improve my farm’s performance	0.319***	0.096	0.369***	0.123
ATTITUDE VARIABLES				
Privacy				
Privacy is important to me	-0.018	0.015	-0.019	0.020
I would be put at a disadvantage if other could access info about my farm	-0.023*	0.014	-0.024	0.018
I feel comfortable sharing information about my farm	0.063***	0.017	0.079***	0.018

Technology use				
I like to have the latest technology	0.005	0.016	0.006	0.020
I find new technologies easy to use	-0.006	0.013	-0.008	0.018
New technology is more hassle than it is worth	-0.024	0.015	-0.033*	0.017
I am getting maximum use out of available tech on my farm	-0.002	0.015	-0.008	0.017
Farm management				
I have implemented new techniques that have been recommended	0.027	0.017	0.025	0.023
I am proactive in seeking advice	0.028*	0.017	0.023	0.023
Precision ag will transform agriculture over the next 20 years	0.01	0.014	0.018	0.017
I know better than others how to manage risk on my farm	-0.029**	0.015	-0.026	0.017
REVENUE RANGE (base = <$100,000)				
$100,000 to $499,999	-0.004	0.067	-0.026	0.097
$500,000 to $999,999	0.042	0.071	0.032	0.100
$1 million to $2 million	0.026	0.075	-0.004	0.102
$2 million to $3 million	0.01	0.093	-0.026	0.126
>$3 million	-0.164**	0.084	-0.206**	0.104
COMPENSATION FOR SURVEY COMPLETION (base = $10)				
$20	-0.011	0.028	-0.001	0.039
$30	-0.009	0.036	-0.008	0.045

*Significance codes: 0.01 '***' 0.05 '**' 0.1 '*'*

DISCUSSION AND CONCLUSION

Farmers may be hesitant to share their data if they feel someone else is using it to make large profits, and the farmers don't see proportional benefits. This could be the case for big data programs operated by for profit groups like crop input suppliers, equipment manufacturers, and financial institutions. Contrarily, most public and private Canadian universities are non-profit organizations, and farmers don't have any direct business relationship with them. These are likely the reasons that contribute to universities being the most popular organization. Cooperating with universities doesn't have any direct negative impact on farmers.

Organizations that farmers have a direct business relationship with are financial institutions, crop input suppliers, and equipment manufacturers. These organizations are profit maximizing. They attempt to capture value from their consumers, in this case farmers. These organizations could use the databank to adjust their pricing strategies, potentially resulting in a loss for farmers. Farmers may be less willing to participate in a big data program run by these organizations because they may be worried about impacts on their business relationships.

Farmers are least likely to share their data with the government. Overapplication of fertilizer, pesticides, and other inputs can have a negative impact on the surrounding environment. Farmers might be scared that if they share their farm level data with the government, it could spark new environmental regulations. Although the data collection process would anonymize farmers in the database, farmers may worry that they could be individually tracked and penalized for certain practices.

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 paper. To make a big data program successful, the farmer participation rate must be high. It's also important to consider cost per respondent. A larger financial compensation increases the participation rate, but increases cost with each respondent. The presence of a non-financial incentive also increases the participation rate. It is up to the individual organization to determine what incentives they can provide. Non-financial incentives can be more difficult to offer because they must be tailored for each respondent. This is possible if a system is put in place to generate them quickly.

The marginal impact of the non-financial incentives is large enough to have an economic impact on response rates, but not as large as the effects of the financial incentives. It would make good business sense for an organization to include these non-financial incentives if they are less costly to generate and administer than the straight cash payment to respondents. There is a large fixed cost component to setting up the non-financial incentive system, but the marginal cost for each additional respondent could be very low. In the ideal case where the data program has many farmers participating, this could result in a very low average cost per respondent.

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 marginal 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 technology oriented individuals from the sample.

The results can loosely be interpreted as farmers trust in the organizations. As farmers trust in an organization increases, they will be more willing to participate in a big data program run by that organization.

REFERENCES

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

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

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

Cox, M., & 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., & 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].

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.

George, G., Haas, M., & Pentland, A. 2004. Big Data and Management. *Academy of Management Journal*, 57(2): 321-326.

Olson, J., Grudin, J., & 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.

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