Abstract
A Monte Carlo simulation study is carried out to examine the effects on study results of subjects completing a survey more than once. Three strategies subjects might use to do this - which is known as farming - are studied. Findings show that farming influences results and can cause both statistical hypothesis testing Type I (false positive) and Type II (false negative) errors in unpredictable ways. A literature review from one management sub-discipline (marketing) was undertaken to investigate how common problem farming might be. Results suggest that while the incentivised survey method which might encourage farming is popular and some approaches to data collection make it difficult to prevent farming altogether, it is unlikely to be commonplace as many research methods prevent it.

Keywords
Survey, Incentive, Online

Authors’ affiliations
1 Nottingham University Business School
2 Edinburgh Napier University

Address for correspondence
thomas.chesney@nottingham.ac.uk Tel: +44 115 846 7759
The impact of repeated lying on survey results

Thomas Chesney* and Kay Penny**

*Nottingham University Business School
**Edinburgh Napier University

October 7, 2010

Abstract

A Monte Carlo simulation study is carried out to examine the effects on study results of subjects completing a survey more than once. Three strategies subjects might use to do this—which is known as farming—are studied. Findings show that farming influences results and can cause both statistical hypothesis testing Type I (false positive) and Type II (false negative) errors in unpredictable ways. A literature review from one management sub-discipline (marketing) was undertaken to investigate how common a problem farming might be. Results suggest that while the incentivised survey method which might encourage farming is popular and some approaches to data collection make it difficult to prevent farming altogether, it is unlikely to be commonplace as many research methods prevent it.

Keywords: survey, incentive, online

This research was sponsored by the Carnegie Trust.

1 Introduction

The convenience for data collection offered by the Internet has proven attractive to researchers. Online methods allow easy access to a wide pool of potential respondents giving data collectors cost effective
global reach (Rhodes et al., 2003). Web based businesses such as SurveyMonkey allow researchers to create and distribute online questionnaires quickly and with low cost, a service that many have used (see http://www.surveymonkey.com/Customer.aspx and http://www.limesurvey.org/en/component/content/article/1-general-news/193-ask-limesurvey-which-universities-are-using-limesurvey). One big disadvantage in this approach is a lack of certainty about the state of respondents during data collection – whether for instance they are tired or inebriated or more fundamentally whether they are who they claim to be. For example, gender swapping in online communities is not uncommon (Turkle, 1995).

Many studies have shown that an incentivised survey often leads to an increased response rate; for example a meta analysis reported by Church (1993) found that the greater the incentive, the greater the increase in response. Thomson et al. (2004) carried out a randomised controlled trial to compare whether one big prize or many small prizes are most effective for encouraging general practitioners to complete a postal survey, and concluded that one big prize increases the response more than many small prizes, despite the lower odds of winning. Bosnjak and Tuten (2003) show that in web surveys, prize draws were more effective in increasing willingness to participate than prepaid or promised monetary incentives.

Repeated lying—also known as farming—is defined as a subject repeatedly completing a survey. Subjects have several potential motivations to do this. Morabia and Zheng (2009) investigate the influence of entry into a raffle as an incentive for participation to an urban transportation survey in New York. Of 3,913 eligible responses, 183 (4.7%) participants were thought to have responded twice as they gave the same email address. The duplicate answers were fairly consistent, although not exactly the same. The authors conclude by saying that surveys involving a raffle should expect multiple entries from the same individual. A less obvious reason is to influence results, which might happen during research into controversial subjects such as employee working conditions. It is also possible that a competing researcher may farm in order to destroy results. Cases of this are unknown but Duffy (2010) demonstrated that it is a real possibility.

The current paper presents and examines three strategies subjects might use for farming, reports a review within one management subdiscipline (marketing) to critically assess the techniques researchers use for dealing with farming, then empirically evaluates what impact
farming might have on research results using a simulation study. We set this work in the context of online surveys, although there is no reason that the results will not apply to any approach to data collection.

2 Review of published papers

Data were collected from the 2008 papers published in ten marketing journals rated highly in the Association of Business Schools journal rankings (see www.the-abs.org.uk). Six journals in the marketing section of the list ranked as world-leading in terms of originality, significance and rigour and four journals ranked as internationally excellent in terms of originality, significance and rigour but which nonetheless falls short of the highest standards of excellence (see: www.rae.ac.uk) were included in the review. The journals were: Journal of Consumer Research, Journal of Marketing, Journal of Marketing Research, Journal of Retailing, Journal of the Academy of Marketing Science, Marketing Science, European Journal of Marketing, Industrial Marketing Management, International Journal of Research in Marketing and International Marketing Review.

In all, 545 papers were examined (Table 1; note that republished papers, editorials, response papers and opinion papers were excluded). Almost one third of all papers reviewed in this study (169 papers, 31.0%) included at least one survey as a data collection method. Twenty-three papers (6.1%) included case studies; 93 (24.7%) reported experiments; 26 (6.9%) were review papers and 105 (27.9%) were theoretical/model development or discussion papers. Of the 169 papers which included a survey, the majority (141 papers) reported on a single survey (83.4%); twenty-one papers (12.4%) included two surveys, and one paper (0.6%) included the results of nineteen surveys. The remaining 3.6% of papers included between three and five surveys.

Of the papers which report the survey method, 47.4% were postal surveys, 23.1% were online surveys, 15.4% were hand delivered, 10.9% were interview surveys and only 3.2% were telephone surveys. A total of 44 surveys (22.4%) included an incentive for the respondent to participate in the survey. The most popular type of incentive was cash (38.6%), followed by a prize draw (20.5%). Three studies offered a gift (6.8%) and two offered refreshments (4.6%). Six studies offered a study report detailing the survey results (13.6%), four offered class
Table 1: The number of papers reviewed in each journal

| Journal                                      | No. of papers reviewed (%) |
|----------------------------------------------|----------------------------|
| Jnl of Consumer Research                     | 76 (13.9%)                 |
| Jnl of Marketing                             | 54 (9.9%)                  |
| Jnl of Marketing Research                    | 54 (9.9%)                  |
| Jnl of Retailing                             | 30 (5.5%)                  |
| Jnl of the Academy of Marketing Science      | 44 (8.1%)                  |
| Marketing Science                            | 58 (10.6%)                 |
| European J of Marketing                      | 74 (13.6%)                 |
| Industrial Marketing Management              | 88 (16.1%)                 |
| International Jnl of Research in Marketing   | 33 (6.1%)                  |
| International Marketing Review               | 34 (6.2%)                  |
| Totals                                       | 545 (100%)                 |

credit to students (9.1%), and three studies did not give any information regarding the type of incentive offered to respondents (6.8%).

There were two common approaches researchers used which counteract farming. (The first approach was almost certainly not used specifically to deal with farming, that is an added benefit of its use.) The first approach involved the researcher being present with the participants when the survey was completed. This does not necessarily mean that the survey was completed on paper; it could be done online in a computer lab with the researcher present. It should be noted that this in itself would not stop a determined farmer who could physically disguise him or herself. However, potential farmers may be unlikely to go to such lengths particularly if they were known to the researcher, or were participating in the survey alongside their fellow peers. Using students as survey participants was a common approach, which also helps in discouraging farming as in many cases each participant would be familiar to the researcher—for example, (Plouffe, 2008) asked students to give their name and email address if they wanted to take part, and (Ofek and Turut, 2008) sent the survey to a group of course participants, although they do not discuss the potential for farming.

The second approach is for the researcher to send the survey to potential respondents and for each survey sent to be identifiable in some way to the researcher—if a survey is sent through the post and contains a unique identifier, then the respondent is unable to photocopy it and return multiple responses. If a hyperlink to a survey is
emailed to participants and each is given a unique token or password, then again they will be unable to complete the survey repeatedly, e.g. Gill (2008). The exception to this would be if the tokens used were sequential and therefore easy for someone to guess, although doing this would likely be discovered by the researcher when a genuine respondent tried to use their token which had been guessed by a farmer. Someone doing this could only expect the ruin data collection, not influence results or gain financially.

None of the papers reviewed explicitly stated that steps were taken to discourage participants from giving multiple responses, but researchers could potentially check for farming after data collection as they will often know the size of their sample and the size of the actual response, the size of a typical expected response and can compare the three. However, this was not always possible according to some of the survey methods used. For example, some papers post requests for participation on bulletin boards or newspapers where the number of people seeing it could only be guessed at (Rick et al., 2008) (Breivik and Thorbjornsen, 2008). Another study used a snowball sampling method to recruit participants with inclusion in a $1000 prize draw as an incentive; although a unique link was emailed to each participant, there remains the potential for farming due to participants recommending their friends as part of the snowball sampling De Bruyn and Lilien (2008).

3 Farming strategies

We have limited behavioural data to indicate exactly how a respondent might farm. Therefore we consider three possible farming strategies that could potentially be used. The first is repeatedly telling the truth. In this case an individual gives their real data more than once. So for example, employing this strategy a 27 year old male would repeatedly claim to be a 27 year old male, and each time accurately report their opinions on the phenomena being researched. The second is answering randomly. Here the 27 year old male may or may not give their data accurately the first time, but on successive completions would randomly invent data. The third technique is where the subject tries to give average responses. So here, if the 27 year old male feels that most respondents will be 18 years old (as might be the case if for instance undergraduate students are completing the survey), he
will invent data in line with this. Of course, he will not be able to
accurately predict what the typical responses will be, but the point
is he will try not to stand out. We have labelled these techniques
*repeated truth, random* and *inlier*.

## 4 Empirical evaluation

A genuine social science dataset \((N = 526)\) was used as the basis for
the empirical evaluation. The data are survey responses on students’
participation and perceptions of e-learning, computer and Internet
use. Only a small subset of variables were included in the empirical
evaluation, and prior to using this subset in the current study, the
data were cleaned and validated, and individuals with impossible or
incorrect data values were excluded, along with individuals who had
missing data values on any of the subset of variables. For the purposes
of this study, an assumption that none of the genuine respondents had
farmed was made.

New records were then generated by Monte Carlo simulation using
the three farming strategies. Three parameters were varied: \(n\) the
size of the sample taken from the dataset of 526 respondents; \(a\) the
proportion of individuals within the sample who farmed; and \(b\) the
number of times they farmed. We examined three levels of each giving
a total combination of 27 experimental conditions. The median \(n\) from
the surveys reviewed in Section 2 was 210 and we used this as our
middle \(n\). For comparison we also analysed datasets of size \(n = 100\)
and \(n = 500\).

A study by Morabia and Tuten (2009) suggests that the proportion
of duplicate entries may be around 5%, but it is unknown how this is
split across \(a\) and \(b\). As data on the frequency or prevalence of farming
do not exist, we looked to known frequencies of plagiarism in students’
coursework (Scanlon and Neuman, 2002) as a guide to set \(a\). These
came from self-reports from students and based on them we set \(a\) to
be 5% and 10%. For the third level we set a low value of \(a = 1\%\). For
\(b\) we used absolute values of 1, 5 and 10 repetitions. We have no data
to support these and they should therefore be considered speculative.

The full survey dataset \((N = 526)\) had already been used to test
a number of hypotheses in an unrelated research project. For the
current study we concentrated on four of its research hypotheses. The
first two were tested by independent samples t-tests; the second two
by one-way ANOVAs:

Hypothesis 1: there is a difference between genders in the average time spent using a computer excluding internet use.

Hypothesis 2: there is a difference between genders in the average time spent using a computer for educational purposes, excluding internet use.

Hypothesis 3: there is a difference between age groups in the average time spent using a computer excluding internet use.

Hypothesis 4: there is a difference between age groups in the average time spent using a computer for educational purposes, excluding internet use.

Results on the full dataset of 526 respondents are presented in Table 2. The distributions of each of the four variables were positively skewed hence, prior to carrying out parametric statistical testing, the data were first transformed using the natural logarithm function in order to ensure that the assumption of normality was met. From testing using all 526 respondents, male students used a computer excluding the Internet for all purposes longer on average than female students ($p = 0.001$). However, there was no significant differences in the time spent using a computer for educational purposes between the genders ($p = 0.676$). The time spent using a computer excluding the Internet differs between age groups for all purposes ($p < 0.001$) and for educational purposes ($p < 0.001$). The summary statistics in Table 2 show that on average, students tend to spend longer using a computer as age increases. Tukey’s multiple comparison tests were calculated to determine which groups differ from each other. Those aged 16 - 20 years use a computer for all purposes significantly less than the older age groups, and those aged 25 years or over use a computer for education purposes significantly more than the two younger age groups.

4.1 Repeated truth

To illustrate the method, consider the second row of Table 3 which shows the combinations of $n$, $a$ and $b$ used in the simulations. For a dataset with 100 responses, 1% of the respondents (the lowest level of $a$) is 1, so our first condition is one respondent farming in a dataset of
Table 2: Natural logarithm of hours spent per week using a computer by full-time students, excluding internet use (n=526) *Independent-samples T-Test **One-way ANOVA

100. A random 100 records were chosen with replacement from the full dataset. The statistical tests were run on this clean data. Then one of the records (our first value of \( b \)) was randomly chosen and copied to create a farmed sample of \( n = 101 \). The statistical tests were run with the farmed sample. We then repeated this with \( b = 5 \) and \( b = 10 \) For our next value of \( a \), 5% (5) random respondents were chosen and the tests were run again for the three levels of \( b \). Finally \( a \) was set to 10% (10). This was repeated 10,000 times, each time taking a fresh sample of \( n = 100 \). The entire process was then repeated with \( n = 210 \) and \( n = 500 \).

### 4.2 Repeated truth results

Tables 4-7 present the proportion of results which were statistically significant at the 5% level for each hypothesis under the repeated truth farming strategy. For Hypotheses 1, 3 and 4 the test results, based on the complete sample of 526 respondents, lead to \( H_0 \) being rejected. For sample sizes of \( n = 100 \) and \( n=210 \), the proportions of significant test results increase slightly as the levels of farming increase. However, for the larger sample size of \( n=500 \), the proportions of significant results decrease slightly as the levels of farming increase.

For Hypothesis 2 (Table 5), where the observed test result based on the complete sample of 526 respondents, is not to reject \( H_0 \), the simulation results demonstrate that the presence of repeated truth farming tends to increase the likelihood of a Type 1 error.
Table 3: Total sample size investigated including farmed cases

| % farmed $(a)$ | Number of repetitions $(b)$ | Sample size $(n)$ |
|----------------|-----------------------------|------------------|
| 0              | 0                           | 100 210 500      |
| 1              | 1                           | 101 212 505      |
| 1              | 5                           | 105 230 525      |
| 1              | 10                          | 110 220 550      |
| 5              | 1                           | 105 215 525      |
| 5              | 5                           | 125 260 625      |
| 5              | 10                          | 150 310 750      |
| 10             | 1                           | 110 231 550      |
| 10             | 5                           | 150 315 750      |
| 10             | 10                          | 200 420 1000     |

Table 4: Proportion of Hypothesis 1 test results which were statistically significant at the 5% level for the 10,000 Repeated Truth Monte Carlo simulated data sets

| Percentage farmed $(a)$ | Number of repetitions $(b)$ | Initial sample size $(n)$ |
|-------------------------|-----------------------------|---------------------------|
|                         | n/a                         | 100 | 210 | 500 |
| 0                       |                             | 0.282 | 0.520 | 0.873 |
| 1                       | 1                           | 0.306 | 0.558 | 0.899 |
| 1                       | 5                           | 0.315 | 0.558 | 0.891 |
| 1                       | 10                          | 0.323 | 0.555 | 0.881 |
| 5                       | 1                           | 0.443 | 0.632 | 0.878 |
| 5                       | 5                           | 0.460 | 0.633 | 0.868 |
| 5                       | 10                          | 0.467 | 0.625 | 0.854 |
| 10                      | 1                           | 0.557 | 0.672 | 0.848 |
| 10                      | 5                           | 0.570 | 0.675 | 0.843 |
| 10                      | 10                          | 0.577 | 0.668 | 0.834 |

Table 4: Proportion of Hypothesis 1 test results which were statistically significant at the 5% level for the 10,000 Repeated Truth Monte Carlo simulated data sets
Table 5: Proportion of Hypothesis 2 test results which were statistically significant at the 5% level for the 10,000 Repeated Truth Monte Carlo simulated data sets

| Percentage farmed (a) | Number of repetitions (b) | Initial sample size (n) | 100 | 210 | 500 |
|-----------------------|---------------------------|-------------------------|-----|-----|-----|
| 0                     | n/a                       | 0.061                   | 0.067 | 0.077 |     |
| 1                     | 1                          | 0.060                   | 0.067 | 0.078 |     |
| 1                     | 5                          | 0.074                   | 0.079 | 0.089 |     |
| 1                     | 10                         | 0.084                   | 0.092 | 0.099 |     |
| 5                     | 1                          | 0.222                   | 0.222 | 0.231 |     |
| 5                     | 5                          | 0.247                   | 0.250 | 0.256 |     |
| 5                     | 10                         | 0.275                   | 0.275 | 0.287 |     |
| 10                    | 1                          | 0.430                   | 0.413 | 0.425 |     |
| 10                    | 5                          | 0.445                   | 0.448 | 0.443 |     |
| 10                    | 10                         | 0.470                   | 0.460 | 0.462 |     |

Table 6: Proportion of Hypothesis 3 test results which were statistically significant at the 5% level for the 10,000 Repeated Truth Monte Carlo simulated data sets

| Percentage farmed (a) | Number of repetitions (b) | Initial sample size (n) | 100 | 210 | 500 |
|-----------------------|---------------------------|-------------------------|-----|-----|-----|
| 0                     | n/a                       | 0.358                   | 0.674 | 0.969 |     |
| 1                     | 1                          | 0.400                   | 0.716 | 0.980 |     |
| 1                     | 5                          | 0.405                   | 0.710 | 0.978 |     |
| 1                     | 10                         | 0.417                   | 0.710 | 0.974 |     |
| 5                     | 1                          | 0.603                   | 0.802 | 0.974 |     |
| 5                     | 5                          | 0.615                   | 0.797 | 0.969 |     |
| 5                     | 10                         | 0.633                   | 0.796 | 0.965 |     |
| 10                    | 1                          | 0.751                   | 0.850 | 0.963 |     |
| 10                    | 5                          | 0.755                   | 0.853 | 0.961 |     |
| 10                    | 10                         | 0.767                   | 0.848 | 0.960 |     |

10
| Percentage farmed (a) | Number of repetitions (b) | Initial sample size (n) |
|----------------------|---------------------------|-------------------------|
| 0                    | n/a                       | 0.577 0.905 1.000       |
| 1                    | 1                         | 0.625 0.929 1.000       |
| 1                    | 5                         | 0.630 0.924 1.000       |
| 1                    | 10                        | 0.634 0.920 1.000       |
| 5                    | 1                         | 0.756 0.940 1.000       |
| 5                    | 5                         | 0.756 0.928 0.999       |
| 5                    | 10                        | 0.764 0.925 0.998       |
| 10                   | 1                         | 0.829 0.939 0.996       |
| 10                   | 5                         | 0.829 0.929 0.996       |
| 10                   | 10                        | 0.835 0.930 0.994       |

Table 7: Proportion of Hypothesis 4 test results which were statistically significant at the 5% level for the 10,000 Repeated Truth Monte Carlo simulated data sets

4.3 Random

The procedure was similar to our strategy for repeated truth with the difference that instead of making a straight copy of $a$ records $b$ times, for each farmed response we randomly generated data between the minimum and maximum legitimate value of each field using a uniform distribution, such that each possible category or value within a particular variable had an equal probability of selection. For gender we randomly selected male or female, for age we randomly generated an age between 16 and 70, and for hours spent using the computer a random number between 0 and 105 (the assumed maximum number of hours possible per day for a week). For hours spent using a computer for education we subtracted a random number between zero and the number generated for total time spent using a computer, from the number generated for total time spent using a computer. (We did this to keep our data within limits so that it would still be sensible, i.e. we did not allow a farmed response to have data where the time spent using a computer for education purposes was greater than the total time spent using a computer, a value which should include the first figure). In the original analysis of the dataset, age was split into three age groups (16-20, 21-24 and 25+). The randomly generated age was recoded into these same groups. (The reason a random age group was not generated was because respondents would not have
Table 8: Proportion of Hypothesis 1 test results which were statistically significant at the 5% level for the 10,000 Random Monte Carlo simulated data sets

| Percentage farmed (a) | Number of repetitions (b) | Initial sample size (n) |
|-----------------------|---------------------------|------------------------|
| 0                     | n/a                       | 0.280 0.524 0.876      |
| 1                     | 1                         | 0.286 0.525 0.875      |
| 1                     | 5                         | 0.285 0.527 0.873      |
| 1                     | 10                        | 0.284 0.524 0.877      |
| 5                     | 1                         | 0.284 0.523 0.875      |
| 5                     | 5                         | 0.283 0.524 0.874      |
| 5                     | 10                        | 0.285 0.526 0.874      |
| 10                    | 1                         | 0.283 0.523 0.874      |
| 10                    | 5                         | 0.283 0.523 0.874      |
| 10                    | 10                        | 0.284 0.524 0.873      |

known that their responses would have been recoded in this way and therefore the existence of age groups could not have influenced farming behaviour.) Notice that given the three levels of \( a \) and \( b \), three of our conditions produce the same amount of random data (\( a = 5, b = 1 \) and \( b = 1, a = 5 \) for instance). In these cases we report both.

### 4.4 Random results

Tables 8-11 present the proportion of results which were statistically significant at the 5% level for each hypothesis under the random farming strategy. The larger the sample size, the higher the proportion of statistically significant results, although the increase is less dramatic for the Hypothesis 2 results. For any particular sample size, there is less variability in the proportion of significant results than the repeated truth results, and increased farming does not appear to have much effect on the results. The exception is for Hypothesis 4 (Table 11); for the larger samples of sizes 210 and 500, increased farming tends to lead to a decrease in the proportion of significant test results i.e. an increase in Type II error.
| Percentage farmed (a) | Number of repetitions (b) | Initial sample size (n) |
|----------------------|---------------------------|------------------------|
|                      |                           | 100        | 210        | 500        |
| 0                    | n/a                       | 0.057      | 0.062      | 0.078      |
| 1                    | 1                         | 0.055      | 0.062      | 0.076      |
| 1                    | 5                         | 0.055      | 0.061      | 0.076      |
| 1                    | 10                        | 0.055      | 0.061      | 0.077      |
| 5                    | 1                         | 0.056      | 0.063      | 0.076      |
| 5                    | 5                         | 0.056      | 0.061      | 0.075      |
| 5                    | 10                        | 0.056      | 0.059      | 0.076      |
| 10                   | 1                         | 0.056      | 0.061      | 0.075      |
| 10                   | 5                         | 0.056      | 0.063      | 0.076      |
| 10                   | 10                        | 0.056      | 0.061      | 0.076      |

Table 9: Proportion of Hypothesis 2 test results which were statistically significant at the 5% level for the 10,000 Random Monte Carlo simulated data sets

| Percentage farmed (a) | Number of repetitions (b) | Initial sample size (n) |
|----------------------|---------------------------|------------------------|
|                      |                           | 100        | 210        | 500        |
| 0                    | n/a                       | 0.358      | 0.667      | 0.971      |
| 1                    | 1                         | 0.388      | 0.681      | 0.969      |
| 1                    | 5                         | 0.386      | 0.683      | 0.968      |
| 1                    | 10                        | 0.389      | 0.678      | 0.968      |
| 5                    | 1                         | 0.384      | 0.677      | 0.967      |
| 5                    | 5                         | 0.385      | 0.682      | 0.966      |
| 5                    | 10                        | 0.386      | 0.681      | 0.967      |
| 10                   | 1                         | 0.389      | 0.682      | 0.967      |
| 10                   | 5                         | 0.390      | 0.684      | 0.967      |
| 10                   | 10                        | 0.388      | 0.683      | 0.968      |

Table 10: Proportion of Hypothesis 3 test results which were statistically significant at the 5% level for the 10,000 Random Monte Carlo simulated data sets
| Percentage farmed (a) | Number of repetitions (b) | Initial sample size (n) |
|----------------------|--------------------------|------------------------|
| 0                    | n/a                      | 0.578 0.903 1.000      |
| 1                    | 1                        | 0.572 0.869 0.972      |
| 1                    | 5                        | 0.568 0.868 0.973      |
| 1                    | 10                       | 0.570 0.867 0.975      |
| 5                    | 1                        | 0.569 0.871 0.973      |
| 5                    | 5                        | 0.567 0.867 0.974      |
| 5                    | 10                       | 0.568 0.869 0.972      |
| 10                   | 1                        | 0.571 0.869 0.974      |
| 10                   | 5                        | 0.571 0.871 0.970      |
| 10                   | 10                       | 0.568 0.871 0.972      |

Table 11: Proportion of Hypothesis 4 test results which were statistically significant at the 5% level for the 10,000 Random Monte Carlo simulated data sets

4.5 Inlier

The same basic procedure as in the previous simulations was used. This time, to produce the inlier farmed records, we used a convenience sample of 38 university students and staff. Each respondent was shown the instructions of the original survey and were asked to guess what they thought the average survey response for each of the four variables (with age group replaced by age) would be. We used these responses as indications of how people would behave when attempting to farm by giving an average response, and inlier records were generated according to the observed probability distributions. The responses from the 38 individuals were used to calculate the probability of someone guessing male; this probability was 0.216. Hence a Bernoulli distribution with a probability of 0.216 was used to generate each farmed case of gender. For respondent age and the two measures of hours using a computer, we used a Normal distribution with mean and standard deviation set according to the survey responses. The distribution used to generate the farmed results for age was N(20.8, 2.22), N(13.6, 8.72) for the average time spent using a computer excluding internet use, and N(9.2, 5.72) for the average time spent using a computer for education purposes only, excluding internet use.
Table 12: Proportion of Hypothesis 1 test results which were statistically significant at the 5% level for the 10,000 Inlier Monte Carlo simulated data sets

| Percentage farmed (a) | Number of repetitions (b) | Initial sample size (n) |
|-----------------------|---------------------------|-------------------------|
|                       |                           | 100 | 210 | 500 |
| 0                     | n/a                       | 0.282 | 0.552 | 0.878 |
| 1                     | 1                         | 0.261 | 0.494 | 0.863 |
| 1                     | 5                         | 0.194 | 0.411 | 0.806 |
| 1                     | 10                        | 0.133 | 0.338 | 0.744 |
| 5                     | 1                         | 0.195 | 0.406 | 0.804 |
| 5                     | 5                         | 0.054 | 0.203 | 0.597 |
| 5                     | 10                        | 0.031 | 0.139 | 0.469 |
| 10                    | 1                         | 0.139 | 0.329 | 0.741 |
| 10                    | 5                         | 0.031 | 0.137 | 0.466 |
| 10                    | 10                        | 0.043 | 0.120 | 0.344 |

5 Inlier results

Tables 12-15 present the proportion of results which were statistically significant at the 5% level for each hypothesis under the inlier farming strategy. As the levels of farming increases, there are dramatic changes in the proportion of statistically significant results. For Hypotheses 1, 3 and 4, the proportions of significant results decrease as farming increases; however, for Hypothesis 2, the proportion of statistically significant results increase.

6 Discussion and Conclusions

The literature review indicated that the incentivised survey method is popular and that some approaches to data collection make it difficult to prevent farming. The simulations demonstrate that farming impacts on results in ways that cannot be predicted. Results are here summarised and then discussed.

In general, the proportion of statistically significant results increases for unfarmed samples with the size of the sample, albeit less dramatically for the Hypothesis 2 results. This increase is not surprising, as an increased sample size will lead to more statistical power, hence the likelihood of a Type II error is decreased.
| Percentage farmed (a) | Number of repetitions (b) | Initial sample size (n) | 100 | 210 | 500 |
|----------------------|---------------------------|-------------------------|-----|-----|-----|
| 0                    | n/a                       |                         | 0.056 | 0.057 | 0.073 |
| 1                    | 1                         |                         | 0.040 | 0.044 | 0.059 |
| 1                    | 5                         |                         | 0.025 | 0.028 | 0.038 |
| 1                    | 10                        |                         | 0.047 | 0.045 | 0.042 |
| 5                    | 1                         |                         | 0.025 | 0.030 | 0.037 |
| 5                    | 5                         |                         | 0.127 | 0.127 | 0.107 |
| 5                    | 10                        |                         | 0.210 | 0.247 | 0.242 |
| 10                   | 1                         |                         | 0.046 | 0.046 | 0.040 |
| 10                   | 5                         |                         | 0.200 | 0.242 | 0.240 |
| 10                   | 10                        |                         | 0.291 | 0.401 | 0.474 |

Table 13: Proportion of Hypothesis 2 test results which were statistically significant at the 5% level for the 10,000 Inlier Monte Carlo simulated data sets

| Percentage farmed (a) | Number of repetitions (b) | Initial sample size (n) | 100 | 210 | 500 |
|----------------------|---------------------------|-------------------------|-----|-----|-----|
| 0                    | n/a                       |                         | 0.365 | 0.684 | 0.968 |
| 1                    | 1                         |                         | 0.343 | 0.646 | 0.948 |
| 1                    | 5                         |                         | 0.265 | 0.539 | 0.873 |
| 1                    | 10                        |                         | 0.194 | 0.426 | 0.779 |
| 5                    | 1                         |                         | 0.259 | 0.535 | 0.869 |
| 5                    | 5                         |                         | 0.092 | 0.238 | 0.588 |
| 5                    | 10                        |                         | 0.052 | 0.116 | 0.408 |
| 10                   | 1                         |                         | 0.196 | 0.427 | 0.789 |
| 10                   | 5                         |                         | 0.053 | 0.122 | 0.401 |
| 10                   | 10                        |                         | 0.046 | 0.076 | 0.248 |

Table 14: Proportion of Hypothesis 3 test results which were statistically significant at the 5% level for the 10,000 Inlier Monte Carlo simulated data sets
Table 15: Proportion of Hypothesis 4 test results which were statistically significant at the 5% level for the 10,000 Inlier Monte Carlo simulated data sets

| Percentage farmed (a) | Number of repetitions (b) | Initial sample size (n) |
|----------------------|---------------------------|-------------------------|
|                      |                           | 100 | 210 | 500 |
| 0                    | n/a                       | 0.582 | 0.907 | 0.999 |
| 1                    | 1                         | 0.359 | 0.737 | 0.994 |
| 1                    | 5                         | 0.113 | 0.363 | 0.913 |
| 1                    | 10                        | 0.074 | 0.197 | 0.743 |
| 5                    | 1                         | 0.115 | 0.361 | 0.909 |
| 5                    | 5                         | 0.050 | 0.092 | 0.365 |
| 5                    | 10                        | 0.047 | 0.060 | 0.179 |
| 10                   | 1                         | 0.073 | 0.192 | 0.733 |
| 10                   | 5                         | 0.049 | 0.059 | 0.173 |
| 10                   | 10                        | 0.048 | 0.049 | 0.085 |

For the repeated truth results, the proportion of significant test results decreases slightly as farming increases for the larger sample size of n=500 for Hypothesis 1, 3 and 4, compared to dramatic increases in the proportions for the smaller samples of 100 and 210 (Tables 4, 6 and 7). This difference is not observed for Hypothesis 2 (Table 5) where the observed data do not lead to a significant test result.

For the random farming method (Tables 8-11), there is less variability in the proportion of significant results for three of the hypotheses. However for Hypothesis 4 (Table 11) increased farming by the random method tends to lead to a decrease in the proportion of significant results for samples of size 210 and 500, thereby showing an increase in Type II error.

For the inlier farming results, the proportion of significant results for Hypotheses 1, 3 and 4 (Tables 12, 14 and 15) decrease dramatically as the levels of farming increase, however, for Hypothesis 2 (Table 13) the proportion of significant results increase in line with an increase in farming.

The repeated truth method of farming may lead to a reduction in variability in the farmed dataset, since repeated responses are added during the farming process. This would lead to an artificial increase in statistical power compared to the unfarmed data, thereby increasing the Type I error. However, these results do not hold for samples of
size \( n = 500 \), where the proportions of statistically significant results
decrease slightly with increased farming.

The random farming method should preserve the variability in
each of the four individual variables however the covariance structure
between the four variables will not be retained since the variables were
randomly generated independently of each other to create the farmed
cases. This most likely explains the reduction in the proportion of
significant results in the larger sample sizes, as the trend in the number
of hours spent using a computer according to gender and age group will
not be retained in the farmed datasets. Even in the smallest sample
sizes where there is little difference in the proportion of significant
results, the trend is there but is not so pronounced.

Although the inlier farming results will lead to a reduction in the
variability of each of the four variables, again the covariance structure
between the farmed cases will not be retained, which explains the
reduction in the proportions of significant results for Hypotheses 1, 3
and 4. Hypothesis 2 results show, on the other hand, a slight increase
in Type I error.

The results presented in this paper show that the effects on sta-
tistical test results depends on which hypothesis is being tested, the
sample size, and the method of farming. Morabia and Zheng (2009)
and Steele et al. (1992) suggest that research results will not be dis-
torted if under 5% of the sample respond twice and on the whole our
data support this. However, the effects on results if a greater propor-
tion of farmers are present, or if farmers were to respond more than
twice have not previously been investigated.

The three types of farming have been simulated to be as realistic
as possible, based on possible techniques which respondents may
use in practice. In reality, some farmers may use more sophisticated
techniques, and further research into the psychological reasons and
strategies of farmers would be beneficial.

It would also be interesting to carry out a simulation study to
explore whether any multivariate statistical techniques have the power
to detect farmed cases, by flagging up any suspect cases which have an
unusual or extreme covariance structure. Any useful methods could
then be recommended as part of a data validity checking procedure
prior to carrying out a statistical analysis of the survey data collected.

The results presented in this paper demonstrate that farming in
online studies is a real and potential problem, and can lead to either
Type I or Type II errors, which cannot be predicted. Farming may
be a considered as a greater issue in some online surveys compared to others, and will be dependent on whether respondents have an incentive to farm. Therefore, when designing an online survey, it is crucial that every possible step is taken to ensure the risk of farming is kept to a minimum, in order to ensure that the data collected are not distorted and that the overall study results are valid and reliable.

References

Bosnjak, M. and Tuten, T.: 2003, Prepaid and promised incentives in web surveys: An experiment, *Social Science Computer Review* **21**(2), 208–217.

Breivik, E. and Thorbjørnsen, H.: 2008, Consumer brand relationships: an investigation of two alternative models, *Journal of the Academy of Marketing Science* **36**(4), 443–472.

Church, A.: 1993, Estimating the effect of incentives on mail survey response rates: A meta-analysis, *Public Opinion Quarterly* **57**, 62–79.

De Bruyn, A. and Lilien, G.: 2008, A multistage model of word-of-mouth influence through viral marketing, *International Journal Of Research in Marketing* **25**, 151–163.

Duffy, J.: 2010, Trust in second life, [http://www.pitt.edu/~jduffy/papers/TrustinSL.pdf](http://www.pitt.edu/~jduffy/papers/TrustinSL.pdf).

Gill, T.: 2008, Convergent products: what functionalities add more value to the base?, *American Marketing Association* **72**, 46–62.

Morabia, A. and Zheng, Y.: 2009, On the influence of a raffle upon responses to an urban transportation survey in new york city, *International Journal of Public Health* **54**, 31–34.

Ofek, E. and Turut, O.: 2008, To innovate or imitate? entry strategy and the role of market research, *Journal of Marketing Research* **XLV**, 575–592.

Plouffe, C.: 2008, Examining “peer-to-peer” (p2p) systems as consumer-to-consumer (c2c) exchange, *European Journal of Marketing* **42**, 1179–1202.
Rhodes, S., Bowle, D. and Hergenrather, K.: 2003, Collecting behavioural data using the world wide web: considerations for researchers, *Journal of Epidemiology and Community Health* **57**, 68–73.

Rick, S., Cryder, C. and Loewenstein, G.: 2008, Tightwads and spendthrifts, *Journal of Consumer Research* **34**(6), 767–782.

Scanlon, P. and Neuman, D.: 2002, Internet plagiarism among college students, *Journal of College Student Development* **43**(3), 374–385.

Steele, T., Schwendig, l. and Kilpatrick, J.: 1992, Duplicate responses to multiple survey mailings: a problem?, *Journal of Advertising Research* **32**(2), 26–34.

Thomson, C., Paterson-Brown, S., Russell, D., McCaldin, D. and Russell, I.: 2004, Encouraging gps to complete postal questionnaires - one big prize or many small prizes? a randomized controlled trial, *Family Practice* **21**(6), 697–698.

Turkle, S.: 1995, *Life on the screen: identity in the age of the Internet*, Simon and Schuster.