Risk Attitudes, Randomization to Treatment, and Self-Selection Into Experiments

by

Glenn W. Harrison, Morten I. Lau and E. Elisabet Rutström *

April 2007

Working Paper 05-01, Department of Economics,
College of Business Administration, University of Central Florida, 2005

Abstract. Randomization to treatment is fundamental to statistical control in the design of experiments. But randomization implies some uncertainty about treatment condition, and individuals differ in their preferences towards taking on risk. Since human subjects often volunteer for experiments, or are allowed to drop out of the experiment at any time if they want to, it is possible that the sample observed in an experiment might be biased because of the risk of randomization. On the other hand, the widespread use of a guaranteed show-up fee that is non-stochastic may generate sample selection biases of the opposite direction, encouraging more risk averse samples into experiments. We undertake a field experiment to directly test these hypotheses that risk attitudes play a role in sample selection. We follow standard procedures in the social sciences to recruit subjects to an experiment in which we measure their attitudes to risk. We exploit the fact that we know certain characteristics of the population sampled, adults in Denmark, allowing a statistical correction for sample selection bias using standard methods. We also utilize the fact that we have a complex sampling design to provide better estimates of the target population. Our results suggest that randomization bias does affect the overall level of risk aversion in the sample we observe, but that it does not affect the demographic mix of risk attitudes in the sample. In complementary laboratory experiments we find additional evidence that the common use of non-stochastic show-up fees generates samples that are more risk averse than would otherwise have been observed.

* Department of Economics, College of Business Administration, University of Central Florida, USA (Harrison and Rutström) and Department of Economics and Finance, Durham Business School, University of Durham, United Kingdom (Lau). E-mail contacts: GHARRISON@RESEARCH.BUS.UCF.EDU, M.I.LAU@DURHAM.AC.UK, and ERUTSTROM@BUS.UCF.EDU.
Harrison and Rutström thank the U.S. National Science Foundation for research support under grants NSF/IIS 9817518, NSF/HSD 0527675 and NSF/SES 0616746. Harrison and Lau thank the Danish Social Science Research Council for research support under project #24-02-0124. Steffen Andersen provided superb research assistance and comments throughout. We are grateful to Jean-Robert Tyran for permission to use the experimental laboratory at the University of Copenhagen. We are also grateful to Shelby Gerking, Richard Hofler and David Scrogin for comments. Supporting data, statistical code and output are stored in the public ExLab Digital Library at http://exlab.bus.ucf.edu.

Randomization to treatment is fundamental to statistical control in the design of experiments. But randomization implies some uncertainty about treatment condition, and individuals differ in their preferences towards taking on risk. Since human subjects often volunteer for experiments, or are allowed to drop out of the experiment at any time if they want to, it is possible that the sample observed in an experiment might be biased because of the risk inherent in randomization.[1] In the extreme case, subjects in experiments might be those that are *least* averse to being exposed to risk. For many experiments of biological response this might not be expected to have any influence on measurement of treatment efficacy, although many laboratory, field and social experiments measure treatment efficacy in ways that could be directly affected by randomization bias.[2] On the other hand, the practice in experimental economics is to offer subjects a fixed participation fee to encourage attendance. These non-stochastic participation fees could offset the effects of randomization, by encouraging *more* risk averse subjects to participate than might otherwise be the case. Thus the term "randomization bias," in the context of economics experiments, should be taken to mean the net effects from these two latent sample selection effects.[3]

We undertake a field experiment and a laboratory experiment to directly test the hypothesis that risk attitudes play a role in sample selection.[4] In both cases we follow standard procedures in the

[1] There is some evidence that other animal species behave as if they have aversion to risk, or prefer risk, in certain environments (Kagel, Battalio and Green [1995; ch.6]). Our focus here is on humans.

[2] Heckman and Smith [1995; p. 99-101] provide many examples, and coin the expression "randomization bias" for this possible effect. Harrison and List [2004] review the differences between laboratory, field, social and natural experiments in economics, and all could be potentially affected by randomization bias. Palfrey and Pevnitskaya [2007] use thought experiments and laboratory experiments to illustrate how risk attitudes can theoretically affect the mix of bidders in sealed-bid auctions with endogenous entry, and thereby change behavior in the sample of bidders observed in the auction.

[3] We hesitate to endorse practices in other fields, in which recruitment fees are not paid to subjects, since they open themselves up to abuse. We have considerable experience of faculty recruiting subjects for "extra credit," but where the task and behavior bears no relationship at all to the learning objectives of the class, and no pedagogic feedback is provided to students even if it does bear some tangential relationship. We have serious ethical problems with such practices, quite apart from the problems of motivation that they raise.

[4] Endogenous subject attrition from the experiment can also be informative about subject preferences, since the subject's exit from the experiment indicates that the subject had made a negative

social sciences to recruit subjects. In our experiments the primary source of randomness has to do with the stochastic determination of final earnings, as explained below. We do also employ random assignment to treatment in some of our experiments, but the general point applies whether the randomness is due to assignment to treatment or random determination of earnings, since the effect is the same on potential subjects. Nevertheless, it is reasonable to suspect that members of most populations from which experimenters recruit participants hold beliefs that the benefits from participating involve some uncertainty. All that is required for sample selection to introduce a bias in the risk attitude of the participants is an expectation of uncertainty, not an actual presence of uncertainty in the experimental task.

In the field experiment we exploit the fact that we know certain characteristics of the population sampled, adults in Denmark in 2003, allowing a correction for sample selection bias using well-known methods from econometrics. The classic problem of sample selection refers to possible recruitment biases, such that the observed sample is generated by a process that depends on the nature of the experiment.[5] In principle, there are two offsetting forces at work in this sample selection process. The use of randomization could attract subjects to experiments that are *less* risk averse than the population, if the subjects rationally anticipate the use of randomization.[6]

---

evaluation of it. See Diggle and Kenward [1994] and Philipson and Hedges [1998] for discussion of this statistical issue.

[5] More precisely, the statistical problem is that there may be some unobserved individual effects that cause subjects to be in the observed sample or not, and these effects could be correlated with responses once in the observed sample. For example, Camerer and Lovallo [1999] find that excess entry into competitive games occurs more often when subjects volunteered to participate knowing that payoffs would depend on skill in a sports or current events trivia. This treatment could encourage less risk averse subjects to participate in the experiment and may explain the observed reference bias effect, or part of it.

[6] It is well known in the field of clinical drug trials that persuading patients to participate in randomized studies is much harder than persuading them to participate in non-randomized studies (e..g., Kramer and Shapiro [1984; p.2742ff.]). The same problem applies to social experiments, as evidenced by the difficulties that can be encountered when recruiting decentralized bureaucracies to administer the random treatment (e.g., Hotz [1992]). For example, Heckman and Robb [1985] note that the refusal rate in one randomized job training program was over 90%.

Conversely, the use of guaranteed financial remuneration, common in experiments in economics for participation, could encourage those that are *more* risk averse to participate.[7]

Our field experiment allows us to evaluate the *net* effect of these opposing forces, which are intrinsic to any experiment in which subjects are voluntarily recruited with financial rewards. We find that measured risk aversion is smaller after we correct for sample selection bias, consistent with the hypothesis that the *use of substantial, guaranteed show-up fees more than offset any bias against attending an experiment that involved randomization*. This effect is statistically significant, implying that there is, in aggregate, a net effect from sample selection due to the opposing influence from randomization and show-up fees on the estimated risk attitudes in our field experiments. We stress, however, that if one had adopted different participation fees there might have been more or less sample selection. We also find that there is no evidence that any sample selection that occurred influenced inferences about the effects of observed individual demographic characteristics on risk aversion.

We then design a laboratory experiment to complement the insights from our field experiment, and explore the conclusion that we might have experienced a larger *gross* sample selection effect due to randomization, but that the muted *net* sample selection effect we observed was due to "lucky" choices of participation fees. Our field design used the same fixed recruitment fee for all subjects, to ensure comparability of subjects in terms of the behavioral task. In the laboratory experiments we exogenously vary this fixed recruitment fee. If the level of the fixed fee affects the risk attitudes of the sample that choose to participate in the experiment, at least over the amounts we consider, we should then be able to directly see different risk attitudes in the sample. As expected *a priori*, we do observe samples that are *more risk averse when we have a higher fixed participation*

[7] Most experiments offer subjects a fixed show-up fee, currently ranging between \$5 and \$10 in convenience samples within the United States. Subjects can also expect to earn an uncertain income, and most experimenters mention this possibility without indicating any expected value or bounds, since that quantitative information could generate biases in the task itself as subject try to attain that earnings threshold.

fee. In another treatment in our laboratory experiments we vary just the *range of the prizes* possible in the task, keeping the fixed participation fee constant. In this case we observe samples that are *more risk averse when we scale the range of prizes up*, compared to the control. Hence, the level of the fixed recruitment fee and information on the range of prizes in the experiment have a direct influence on the composition of the sample in terms of individual risk attitudes.

The implication is that experimental economists should pay much more attention to the process that leads subjects to participate in the experiment if they are to draw reliable inferences in any setting in which risk attitudes play a role. This is true whether one conducts experiments in the laboratory or the field. We consider the role of preferences over risk, which is central to the experimental method because of the role of randomization to treatment. But the same concerns apply to the elicitation of other types of preferences, such as social preferences or time preferences. These concerns arise when subjects have some reason to believe that the task will lead them to evaluate those preferences, such as in longitudinal designs allowing attrition, or social experiments requiring disclosure of the nature of the task prior to participation. They might also arise if the sample is selected by some endogenous process in which selection might be correlated with those preferences, such as group membership or location choices.

1. Data

A. The Task

We employ a simple experimental measure for risk aversion introduced by Holt and Laury [2002] and extended by Harrison, Lau, Rutström and Sullivan [2005]. Each subject is presented with a series of choices between two lotteries, which we call A or B. All choices are presented simultaneously to the subject. In the field experiments the first choice involves a 10% chance of receiving DKK2,000 and a 90% chance of receiving DKK1,600. The expected value of this lottery is DKK1,640. Lottery B in the first choice gives a 10% chance of receiving DKK3,850 and 90% chance of receiving DKK100, for an expected value of DKK480. Thus the two lotteries have a relatively large difference in expected values, in this case DKK1,170. As one proceeds down the payoff table the probability of winning the high prize in each lottery increases, and the expected value of lottery B steadily becomes greater than the expected value of lottery A.

Subjects are typically confronted with ten such choices. The subject chooses A or B in each row, and one row is later selected at random for payout for that subject. The logic behind this test for risk aversion is that only risk-loving subjects would take lottery B in the first choice presented above, and only risk-averse subjects would take lottery A in the last few choices. A subject that is neutral to risk should switch from choosing A to B when the EV of each is about the same, so a risk-neutral subject would choose A for the first four choices and then choose B thereafter in a payoff table with 10 symmetric intervals. In addition to the A/B choice on each row there is also an option to express indifference.

These data may be analyzed using a constant relative risk aversion (CRRA) utility characterization of utility. The CRRA utility of each lottery prize y is defined as $U(y)=y^{1-r}/(1-r)$, where r is the CRRA coefficient.[8] For each row one can calculate the implied bounds on the CRRA coefficient, and these intervals are reported by Holt and Laury [2002, Table 3].

This basic task is extended in several ways in the field experiments. First, each subject is

---

[8] With this parameterization, $r = 0$ denotes risk neutral behavior, $r > 0$ denotes risk aversion, and $r < 0$ denotes risk loving. When $r = 1$, $U(m) = \ln(m)$. Harrison, Lau and Rutström [2007] demonstrate that CRRA is an appropriate functional form for the adult Danish population, by nesting it into more general functional forms which they estimate and testing the CRRA restriction directly. There is also evidence that risk attitudes elicited in this manner are temporally stable: see Harrison, Johnson, McInnes and Rutström [2005a]. We evaluate risk attitudes assuming a conventional expected utility analysis characterization of choice under behavior, but obviously one could substitute a non-expected utility characterization. We would not expect any differences in our main conclusions if this substitution was made.

presented with four such lottery pairs, so that CRRA is elicited for the same subject over a wider range of income levels.[9] Thus we have repeat measures for each subject, and use appropriate statistical models for such data. Second, subjects were randomly assigned to treatments designed to test if their risk attitudes were affected by the way that the task was "framed," since one might expect that some subjects would simply choose to switch in the middle of a series of lotteries.[10] Third, we "iterate" the choices made by each subject so that we refine the interval at which they switch from A to B.[11] This allows us to focus on responses that consist of small CRRA intervals. In fact, we will reduce these responses to their mid-points, and view the subject as providing a point response, since the intervals are so small at the end of this iterative process.[12]

We ask the subject to respond to all four risk aversion tasks and then randomly decide which one to play out. The large incentives and budget constraints precluded us from paying all subjects, so each subject is given a 10% chance to actually receive the payment associated with his decision.

---

[9] The four sets of prizes are as follows, with the two prizes for lottery A listed first and the two prizes for lottery B listed next: (A1: 2000 DKK, 1600 DKK; B1: 3850 DKK, 100 DKK), (A2: 2250 DKK, 1500 DKK; B2: 4000 DKK, 500 DKK), (A3: 2000 DKK, 1750 DKK; B3: 4000 DKK, 150 DKK), and (A4: 2500 DKK, 1000 DKK; B4: 4500 DKK, 50 DKK). At the time of the field experiments, the exchange rate was approximately 6.55 DKK per U.S. dollar, so the prizes range from approximately $7.65 to $687.

[10] We devise a test for framing effects by varying the cardinal scale of the multiple price list. Two asymmetric frames are developed: the *skewHI* treatment offers initial probabilities of (0.3, 0.5, 0.7, 0.8, 0.9, 1.0) while the *skewLO* treatment offers initial probabilities of (0.1, 0.2, 0.3, 0.5, 0.7, 1.0). These two asymmetric treatments yield 6 decision rows in the first payoff table of each task, as opposed to 10 rows in the symmetric frame.

[11] We extend the multiple price list (MPL) to allow more refined elicitation of the true risk attitude. The basic MPL is the standard format in which the subject sees a fixed array of paired options and chooses one for each row. The iterative iMPL format extends this by first asking the subject to simply choose the row at which he wants to first switch from option A to option B, assuming monotonicity of the underlying preferences to automatically fill out the remaining choices. The second step is to then allow the individual to make choices from refined options within the option last chosen. That is, if someone decides at the first stage to switch from option A to option B between probability values of 0.1 and 0.2, the second stage of an iMPL would then prompt the subject to make more choices *within* this interval, to refine the values elicited. The comparative properties of the iMPL and MPL institutions are studied in Andersen, Harrison, Lau and Rutström [2006].

[12] The distribution of the elicited CRRA interval size is right skewed with a mean of 0.17 and a median of 0.03. More than 25% of the observations are point estimates and the inter-quartile range is between 0 and 0.09.

B. Field Sampling Procedures

The sample for the field experiments was designed to be representative of the adult Danish population in 2003. There were six steps in the construction of the sample, detailed in Harrison, Lau, Rutström and Sullivan [2005] and essentially following those employed in Harrison, Lau and Williams [2002]:

- First, a random sample of 25,000 Danes was drawn from the Danish Civil Registration Office in January 2003. Only Danes born between 1927 and 1983 were included, thereby restricting the age range of the target population to between 19 and 75. For each person in this random sample we had access to their name, address, county, municipality, birth date, and sex. Due to the absence of names and/or addresses, 28 of these records were discarded.
- Second, we discarded 17 municipalities (including one county) from the population, due to them being located in extraordinarily remote locations, and hence being very costly to recruit. The population represented in these locations amounts to less than 2% of the Danish population, or 493 individuals in our sample of 25,000 from the Civil Registry. Hence it is unlikely that this exclusion could quantitatively influence our results on sample selection bias.
- Third, we assigned each county either 1 session or 2 sessions, in rough proportionality to the population of the county. In total we assigned 20 sessions. Each session consisted of two sub-sessions at the same locale and date, one at 5pm and another at 8pm, and subjects were allowed to choose which sub-session suited them best.
- Fourth, we divided 6 counties into two sub-groups because the distance between some municipalities in the county and the location of the session would be too large. A weighted random draw was made between the two sub-groups and the location selected, where the

weights reflect the relative size of the population in September 2002.

- Fifth, we picked the first 30 or 60 randomly sorted records within each county, depending on the number of sessions allocated to that county. This provided a sub-sample of 600.
- Sixth, we mailed invitations to attend a session to the sub-sample of 600, offering each person a choice of times for the session. Response rates were low in some counties, so another 64 invitations were mailed out in these counties to newly drawn subjects.[13] Everyone that gave a positive response was assigned to a session, and our recruited sample was 268.

Attendance at the experimental sessions was extraordinarily high, including 4 persons who did not respond to the letter of invitation but showed up unexpectedly and participated in the experiment. Four persons turned up for their session, but were not able to participate in the experiments.[14] These experiments were conducted in June of 2003, and a total of 253 subjects participated.[15] Sample weights for the subjects in the experiment can be constructed using this experimental design, and can be used to calculate weighted distributions and averages that better reflect the adult population of Denmark.

The initial recruitment letter for the field experiments explained the purpose of the experiment and that it was being conducted by the Ministry of Economic and Business Affairs of the Danish Government. The letter clearly identified that there would be some randomization

[13] We control for the recruitment wave to which the subject responded in our statistical analysis of sample selection.

[14] The first person suffered from dementia and could not remember the instructions; the second person was a 76 year old woman who was not able to control the mouse and eventually gave up; the third person had just won a world championship in sailing and was too busy with media interviews to stay for two hours; and the fourth person was sent home because they arrived after the instructions had begun and we had already included one unexpected "walk-in" to fill their position.

[15] Certain events might have plausibly triggered some of the no-shows: for example, 3 men did not turn up on June 11, 2003, but that was the night that the Danish national soccer team played a qualifying game for the European championships against Luxembourg that was not scheduled when we picked session dates.

involved in determining earnings. In translation, the uncertainty was explained as follows:

> **You can win a significant amount**
> To cover travel costs, you will receive 500 kroner at the end of the meeting. Moreover, each participant will have a 10 percent chance of receiving an amount between 50 and 4,500 kroner in the first part of the survey, and this amount will also be paid at the end of the meeting. In the second part of the survey, each participant will have a 10 percent chance of receiving at least 3,000 kroner. A random choice will decide who wins the money in both parts of the survey.

Of course, this paragraph also made it clear that there would be a fixed, non-stochastic income of 500 kroner. The earnings referred to as the "first part of the survey" were for the risk aversion task, and the earnings referred to as the "second part of the survey" were for a separate task eliciting individual discount rates. Thus we know that subjects should have rationally anticipated the use of randomization in these experiments.

C. Laboratory Sampling Procedures

The initial set of laboratory experiments were conducted in October 2003. We recruited 100 subjects from the University of Copenhagen and the Copenhagen Business School. All subjects were recruited using the *ExLab* software.[16] The sessions were announced in 7 different lectures. At each lecture an announcement of the experiment was read aloud, and subjects were asked to enroll for the experiment by accessing *ExLab* through the Danish web page for this project. Of the 100 subjects recruited, 90 showed up for the experiment evenly spread across the 9 sessions.

Before the subjects signed up for one of the nine sessions, the web page provided them with the following information about payments:

> You will be paid 250 kroner at the end of the meeting, and you can earn an additional considerable sum of money. Each participant will have a 10 percent chance of receiving an

[16] This recruitment software is available for academic use at http://exlab.bus.ucf.edu. In addition, all instructions for our experiments are provided for public review at the *ExLab* Digital Library at the same location.

> amount between 50 kroner and 4,500 kroner in the first part of the survey, and this amount will also be paid at the end of the meeting. In the second part of the survey, each participant will have a chance of receiving at least 3,000 kroner. A random choice will decide who wins the money in both parts of the survey.

The subjects recruited for the laboratory experiments were thus given the same information about payments as the field subjects, but the fixed recruitment fee in the laboratory experiments is reduced to DKK250.

The second set of laboratory experiments was conducted in November 2006 in Copenhagen. The sessions were again announced at numerous lectures and subjects were asked to enroll for the experiment by accessing *ExLab*. The subjects were randomly split across two recruitment treatments. Compared to the control group in the initial set of laboratory experiments, the first treatment reduced the fixed recruitment fee to DKK100, while in the second treatment all prizes in the experiment were scaled down by 50%. Letters of invitation were sent out by email to all subjects, and they were provided with the same information as the control group in the initial set of laboratory experiments, except for the obvious changes in the fixed and variable payments. Subjects were informed that a maximum of 20 people could participate in the meeting, and they were signed up in the order they replied to the email. We had 15 subjects who were recruited with the lower fixed participation fee, and 20 subjects were recruited with the lower range of prizes. This provides us with an overall sample of 125 subjects from the two sets of laboratory experiments.

D. Conduct of the Experiment

The experiment was conducted in four parts. Part I consisted of a questionnaire collecting subjects' socio-demographic characteristics. Specifically, we collected information on age, gender, size of town the subject resided in, type of residence, primary occupation during the last 12 months, highest level of education, household type (viz., marital status and presence of younger or older children), number of people employed in the household, total household income before taxes, whether the subject is a smoker, and the number of cigarettes smoked per day. Part IV consisted of another questionnaire which elicits information on the subject's financial market instruments, and probes the subject for information on their expectations about their future economic conditions and their own future financial position. The questionnaires are rather long, so we chose to divide them across Parts I and IV in order to reduce subject fatigue and boredom. Part II consisted of the four risk aversion tasks, and Part III presented subjects with four or six individual discount rate tasks similar to those developed in Harrison, Lau and Williams [2002]. We will not discuss the individual discount rate findings here.[17]

The four risk aversion tasks incorporate the incentive structure and assigned frames described earlier. After subjects completed the four tasks, several random outcomes were generated in order to determine subject payments. For all subjects, one of the four tasks was chosen, then one of the decision rows in that task was chosen. To maintain anonymity we performed the draws without announcing to which subjects it would apply. In the case where a subject indicated indifference for the chosen decision row, another random draw determined whether the subject received the results from Lottery A or Lottery B. At this point all subjects knew whether they were playing Lottery A or Lottery B, and another random draw determined whether subjects were to receive the high payment or the low payment. Finally, a 10-sided die was rolled for each subject. Any subject who received a roll of "0" received actual payment according to that final outcome. All payments were made at the end of the experiment. A significant amount of time was spent training subjects on the choice tasks and the randomization procedures in Part II of the experiment.

[17] If there is some effect of sample selection on risk attitudes, then one would expect to see a direct effect on inferred discount rates, which equalize the present value of discounted utility streams (Andersen, Harrison, Lau and Rutström [2005]). Additional treatments in the initial set of laboratory experiments are reported in Andersen, Harrison, Lau and Rutström [2006].

3. Results

A. Field Experiments

In order to assess the importance of sample selection on risk attitudes in the field experiment, we applied regression models that condition on observable characteristics of the subjects and allow for selection biases using techniques standard in econometrics.[18] Table 1 provides the definitions of the explanatory variables and summary statistics. Table 2 displays the results from maximum-likelihood estimation of a sample selection model of elicited risk attitudes, as well as a comparable model that does not allow for sample selection. Both sets of estimates allow for the complex survey design. In particular, they adjust estimates for the fact that subjects in one county were selected independently of subjects in other counties, that sample weights for each subject reflect the adult population of Denmark, as well as the possibility of correlation between responses by the same subject.[19]

The results indicate that the sample estimates of the main CRRA equation are reliable conditional on the characteristics of the sample observed in the experiment, but that there was evidence of significant sample selection into the experiment. The ancillary parameter ρ measures the estimated correlation between the residuals of the sample selection equation and the main CRRA

---

[18] See Vella [1998] for a review of the range of techniques available. We employed full information maximum likelihood estimation of the parametric Heckman [1976][1979] selection model, with corrections to standard errors for the complex sample survey design employed. Version 8.2 of *Stata* was employed to undertake the estimation: see StataCorp [2003] for documentation.

[19] The use of clustering to allow for "panel effects" from unobserved individual effects is common in the statistical survey literature. Clustering commonly arises in national field surveys from the fact that physically proximate households are often sampled to save time and money, but it can also arise from more homely sampling procedures. For example, Williams [2000; p.645] notes that it could arise from dental studies that "collect data on each tooth surface for each of several teeth from a set of patients" or "repeated measurements or recurrent events observed on the same person." The procedures for allowing for clustering allow heteroskedasticity between and within clusters, as well as autocorrelation within clusters. They are closely related to the "generalized estimating equations" approach to panel estimation in epidemiology (see Liang and Zeger [1986]), and generalize the "robust standard errors" approach popular in econometrics (see Rogers [1993]).

equation. It equals 0.46, has a standard error of only 0.18, and has a 95% confidence interval with values of +0.04 and +0.96. If this correlation had been zero then there would have been no evidence of sample selection bias on the main estimates of CRRA. The coefficients in the sample selection are jointly significant, as are many of the individual coefficients. On the other hand, the coefficient estimates for the main CRRA equation are virtually identical.[20]

To estimate the net effects on estimated risk aversion of allowing for sample selection, we re-estimate the specification in Table 1 using only the task characteristics for the main CRRA equation. In this case the constant term picks up the joint effect of all of the demographic effects; the constant term in Table 1 only reflects the default individual (i.e., the one for whom all of the dummy variables take on the value 0), rather than a representative individual. With no sample selection correction we estimate CRRA to be 0.45, with a 95% confidence interval between 0.31 and 0.59; with the sample selection correction we estimate CRRA to be 0.23 with a 95% confidence interval between 0.02 and 0.45. These latter estimate is significantly different from 0.45, with a *p*-value of 0.048. Thus we have *evidence that our sample is more risk averse than the population*, where the population is interpreted as the estimate correcting for sample selection. Thus the non-stochastic show-up fee we used appears to have been "overkill" on average, more than offsetting the effects of expected randomization in the experiment.

B. Laboratory Experiments

The laboratory experiments allow us to test the effect of variations in the recruitment information on individual risk attitudes, since the information was exogenous and provided at

[20] Formal tests of pairwise equality for the estimates in the two specifications support this conclusion at any standard significance level. Similar results are obtained for a test of the joint hypothesis that all coefficient estimates in the two specifications are the same.

random to subjects. Under the maintained hypothesis that the individual risk attitudes are not affected by the recruitment fee we can directly test the impact on the sample composition by estimating a shift in risk attitudes from the announced change in the payment conditions. Table 3 reports results from an interval regression model of elicited CRRA values from Task 4 in our laboratory experiments, controlling for recruitment treatments, framing conditions, task effects, experimenter and individual characteristics.[21] We estimate the model independently for each task but report only the task for which we see significant treatment effects. The regressions for the other tasks are qualitatively similar, but with lower significance levels. The effect from the fixed recruitment fee is positive, as expected, and significant using a one-tailed test. We find that there is a significant effect from using a lower fixed recruitment fee. The coefficient is equal to 0.34 with a one-tailed *p*-value of 0.045 and a 95% confidence interval with values of −0.05 and 0.73. As expected *a priori*, using a *higher* fixed recruitment fee results in a sample that includes subjects with *greater* aversion to risk, since they self-select not to attend when the fixed recruitment fee is lower.

We also find a significant effect from changing the range of lottery prizes. The coefficient is 0.36, has a one-tailed *p*-value of 0.010 and a 95% confidence interval with values of 0.06 and 0.66. This effect is consistent with the hypothesized sample selection effect, but could of course also be due to relative risk attitudes being increasing rather than constant over the prize domain. Harrison, Lau and Rutström [2007] present evidence that Danes exhibit CRRA over the same prize domain as the high one used here, lending support for the assumption of CRRA. Nevertheless, our treatment

---

[21] We used a multiple price list with 10 symmetric intervals in the second set of laboratory experiments, and responses are coded as using the upper and lower boundaries of the elicited CRRA intervals. We do not report a pooled regression here since a Hausman test of the random-effects specification indicates that it is not valid, and we would not be able to test for treatment effects using fixed-effect models since our treatments are between subject. Nevertheless, pooled estimations of a random effects specification show the same qualitative result. This suggests that the specification error detected by the Hausman test perhaps pertains to some other aspect of the pooled model than the one we focus on.

amounts to a scaling up of the lottery prizes presented to subjects, which is qualitatively the same design that Holt and Laury [2002] used to show evidence of increasing RRA (IRRA) in college student subjects in the United States.[22] We conclude that both treatment effects point in the same direction: increasing the fixed or random payments may generate a sample that is more risk averse than we would otherwise observe.

Figure 1 displays the kernel distribution of predicted risk attitudes across the two fixed participation fee treatments. The predictions are based on the demographic samples in the two treatment groups, so they do not correct for sample composition differences. We find that the mean CRRA value in Task 4 is 0.81 for the high fixed fee and 0.67 for the low fixed fee, and the difference between the two values is significant at conventional levels. Since the distributions capture heterogeneity in preferences we can conclude that preference heterogeneity is not large enough to mask the shift in the mean.

From Figure 2 we see the kernel distribution of predicted risk attitudes across the two prize range treatments. Consistent with the marginal effects observed in Table 3, we find that the mean CRRA value is significantly higher for the high lottery prizes compared to the low ones. The mean CRRA for the high prize treatment is 0.81 and for the low it is 0.59 and this difference is significant.

4. Conclusions

Heckman and Smith [1995; p.99] noted that, "Surprisingly, little is known about the empirical importance of randomization bias." Aggregative data about participation rates from job training experiments by Hotz [1992] and clinical trials by Kramer and Shapiro [1984] suggest that it

[22] Harrison, Johnson, McInnes and Rutström [2005b] and Holt and Laury [2005] show that their evidence of IRRA is robust to the confound of an order effect, even though the quantitative extent of IRRA is smaller once one controls for that counfound.

might be significant, but we know of no study that directly evaluates the hypothesis.

Our results suggest that randomization bias does have an effect in our field experiments, given the fixed participation fee we used. We present evidence that *the use of non-stochastic show-up fees, relatively standard in experimental economics, may have generated a sample that was more risk averse than would otherwise have been observed.* Thus, one needs to pay special attention to the expectations of earnings induced on participants during the recruitment. In general, one should always correct for possible sample selection biases, but in the case of risk aversion and standard experimental practices regarding recruitment, there is perhaps a stronger case to be made for such corrections.

The recruitment procedures for the field and laboratory experiments were typical of those used in standard economics experiments, in the sense that they referred to a fixed participation fee which could have offset the selection effects of randomization. We also conducted a complementary treatment to compare the effects of varying the random component of participation rewards, namely a scaling of the range of the lottery prizes, to determine if that influences the risk attitudes of the observed sample.[23] The results from our laboratory experiments suggest that information on the range of possible prizes appears to generate the same kind of self-selection as the fixed participation fee. However, risk attitudes need not be the same for all prizes or outcomes. In experiments that employ treatments with very high prizes (e.g., life or death in the case of medical interventions), then our results may need to be modified to also reflect the possibility of increasing relative risk aversion.

There are qualifications and possible extensions to our analysis from the field experiments.

[23] Rutström [1998] undertakes a design of this kind in the laboratory, varying the show-up fee. She finds considerable differences in the individual characteristics of the subjects that turn up as one varies that fee, but does not conduct a formal test of the effects on elicited risk attitudes. Lazear, Malmendier and Weber [2004] design an experiment to test if subjects recruited into a session endogenously sort away from a task that would involve the expression of social preferences, and report significant evidence that they do. Thus their design embeds one sample selection step within an overall experiment, allowing it to be studied intensively. They do not consider sample selection into the overall experiment, or obviously the effects on risk attitudes since that was not an objective of their experiment.

First, one would always like to know "more" about the population being sampled. The Danish environment is a relatively rich one, in which we could identify three characteristics of the subjects before observing if they agreed to participate. In many other cases less is known about those who are not participating and there is always a risk that the statistical model of participation is poorly specified. Since the subjects that do not participate are, by their nature, unobserved from the perspective of the experimenter, this problem is likely to be a general one. Second, one could build into the experimental design even tighter tests for possible sample selection effects due to randomization bias. It would be possible to vary the fixed and random participation fees used for recruitment purposes to ensure that the expected value of rewards are the same across treatments. By varying the mix of rewards in terms of the random and non-random component, but ensuring that the expected value remain the same, this could provide a more finely calibrated experimental design to detect sample selection due to randomization bias.[24] Finally, there is now a rich econometric literature on less parametric specifications of corrections for sample selection. Given the known importance of parametric structure in the standard sample selection specifications we employ, it would be valuable future research to investigate the robustness of our results to the use of those specifications.[25]

[24] Or one could calculate designs that use mean-preserving spreads of random rewards around the expected *utility* of rewards, if one had some null hypothesis as to the risk aversion of subjects.

[25] For example, see Das, Newey and Vella [2003] and their references to the literature.

Table 1: List of Variables and Descriptive Statistics for Field Experiments

Variable	Definition	Estimated Population Mean	Raw Sample Mean
female	Female	0.50	0.51
young	Aged less than 30	0.19	0.17
middle	Aged between 40 and 50	0.27	0.28
old	Aged over 50	0.33	0.38
single	Lives alone	0.21	0.20
kids	Has children	0.31	0.28
nhhd	Number of people in the household	2.54	2.49
owner	Owns own home or apartment	0.68	0.69
retired	Retired	0.13	0.16
student	Student	0.10	0.09
skilled	Some post-secondary education	0.38	0.38
longedu	Substantial higher education	0.36	0.36
IncLow	Lower level income	0.33	0.34
IncHigh	Higher level income	0.36	0.34
copen	Lives in greater Copenhagen area	0.27	0.27
city	Lives in larger city of 20,000 or more	0.41	0.39
experimenter	Experimenter Andersen (default is Lau)	0.47	0.49

Legend: Most variables have self-evident definitions. The omitted age group is 30-39. Variable "skilled" indicates if the subject has completed vocational education and training or "short-cycle" higher education, and variable "longedu" indicates the completion of "medium-cycle" higher education or "long-cycle" higher education. These terms for the cycle of education are commonly used by Danes (most short-cycle higher education program last for less than 2 years; medium-cycle higher education lasts 3 to 4 years, and includes training for occupations such as a journalist, primary and lower secondary school teacher, nursery and kindergarten teacher, and ordinary nurse; long-cycle higher education typically lasts 5 years and is offered at Denmark's five ordinary universities, at the business schools and various other institutions such as the Technical University of Denmark, the schools of the Royal Danish Academy of Fine Arts, the Academies of Music, the Schools of Architecture and the Royal Danish School of Pharmacy). Lower incomes are defined in variable "IncLow" by a household income in 2002 below 300,000 kroner. Higher incomes are defined in variable "IncHigh" by a household income of 500,000 kroner or more.

Table 2: Estimated Relative Risk Aversion in Field Experiments

Maximum likelihood estimates, with standard errors corrected for complex survey design

		Sample Selection Correction			No Correction		
Variable	Variable Description	Estimate	Standard Error	*p*-value	Estimate	Standard Error	*p*-value
		A. CRRA Equation					
Constant		-0.10	0.24	0.68	0.08	0.23	0.72
skewLO	Frame to skew RA down	-0.18	0.10	0.08	-0.17	0.10	0.08
skewHI	Frame to skew RA up	0.29	0.08	0.00	0.30	0.08	0.00
Task2	Second RA task	0.28	0.06	0.00	0.29	0.06	0.00
Task3	Third RA task	0.22	0.05	0.00	0.22	0.05	0.00
Task4	Fourth RA task	0.18	0.04	0.00	0.17	0.04	0.00
experimenter	Experimenter Steffen Andersen	-0.05	0.08	0.47	-0.03	0.08	0.74
female	Female	0.01	0.07	0.89	0.03	0.07	0.64
young	Aged less than 30	0.15	0.17	0.36	0.13	0.17	0.45
middle	Aged between 40 and 50	-0.29	0.12	0.01	-0.32	0.12	0.01
old	Aged over 50	-0.19	0.13	0.15	-0.19	0.14	0.17
single	Lives alone	0.14	0.12	0.24	0.15	0.12	0.22
kids	Has children	-0.02	0.11	0.83	-0.03	0.11	0.76
nhhd	Number in household	0.02	0.05	0.71	0.02	0.05	0.70
owner	Own home or apartment	0.18	0.09	0.06	0.17	0.09	0.07
retired	Retired	0.03	0.11	0.76	0.03	0.11	0.80
student	Student	0.27	0.14	0.05	0.27	0.14	0.06
skilled	Some post-secondary education	0.27	0.09	0.00	0.27	0.09	0.00
longedu	Substantial higher education	0.34	0.10	0.00	0.35	0.10	0.00
IncLow	Lower level income	-0.02	0.10	0.84	-0.03	0.10	0.80
IncHigh	Higher level income	0.01	0.10	0.94	0.01	0.10	0.94
copen	Lives in Copenhagen area	0.12	0.10	0.23	0.08	0.10	0.42
city	Lives in larger city of 20,000 or more	0.06	0.09	0.48	0.05	0.09	0.57
		B. Sample Selection Equation					
Constant		0.75	0.10	0.00			
female	Female	-0.14	0.09	0.14			
young	Aged less than 30	0.13	0.14	0.34			
middle	Aged between 40 and 50	0.22	0.13	0.09			
old	Aged over 50	0.01	0.12	0.96			
County_15	County 15	-0.24	0.08	0.00			
County_20	County 20	-0.35	0.09	0.00			
County_25	County 25	-0.41	0.11	0.00			
County_30	County 30	-0.58	0.09	0.00			
County_42	County 42	-0.30	0.07	0.00			
County_50	County 50	-0.42	0.11	0.00			
County_55	County 55	-0.52	0.13	0.00			
County_60	County 60	0.03	0.09	0.71			
County_65	County 65	-0.05	0.09	0.57			
County_70	County 70	-0.32	0.08	0.00			
County_80	County 80	-0.40	0.09	0.00			
wave2	Second wave of invitations	-0.39	0.23	0.09			
wave3	Third wave of invitations	-0.07	0.39	0.86			
ρ	Error correlation	0.46	0.18				
σ	Standard error of residual in CRRA equation	0.74	0.04				

Table 3: Estimated Relative Risk Aversion in Laboratory Experiments

Interval regression of Task 4

Variable	Description	Estimate	Standard Error	*p*-value	Lower 95% Confidence Interval	Upper 95% Confidence Interval
Constant		0.49	0.34	0.15	-0.17	1.16
fixed_fee	High fixed participation fee used	0.34	0.20	0.09	-0.05	0.73
variable_earn	High variable earnings used	0.36	0.15	0.02	0.06	0.66
skewLO	Frame to skew RA down	-0.22	0.15	0.13	-0.51	0.06
skewHI	Frame to skew RA up	-0.01	0.13	0.93	-0.26	0.24
experimenter	Experimenter Steffen Andersen	-0.12	0.12	0.30	-0.35	0.11
female	Female	0.08	0.11	0.48	-0.14	0.29
single	Lives alone	-0.31	0.14	0.02	-0.58	-0.04
nhhd	Number in household	0.01	0.09	0.93	-0.16	0.18
owner	Own home or apartment	-0.12	0.15	0.42	-0.42	0.18
student	Student	-0.11	0.14	0.41	-0.38	0.15
skilled	Some post-secondary education	-0.12	0.10	0.24	-0.32	0.08
IncLow	Lower level income	0.09	0.19	0.62	-0.27	0.46
IncHigh	Higher level income	-0.07	0.22	0.75	-0.50	0.36
copen	Lives in Copenhagen area	-0.04	0.15	0.78	-0.34	0.26

Figure 1: Effect of Varying the Fixed Participation Fee

Predicted CRRA for Danish Lab Samples

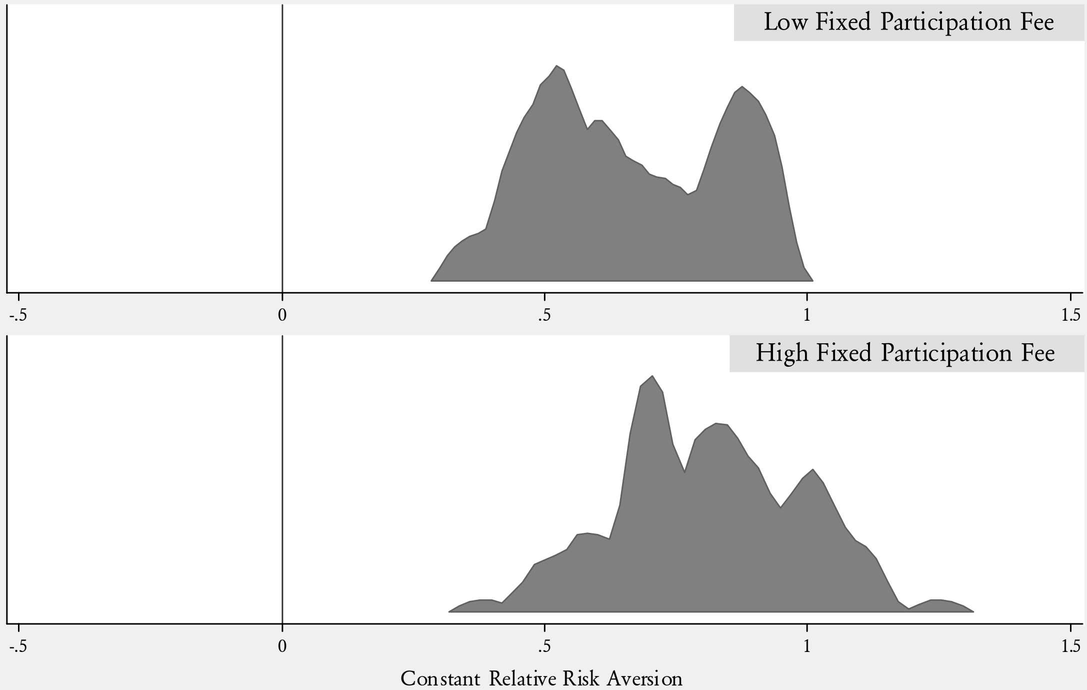
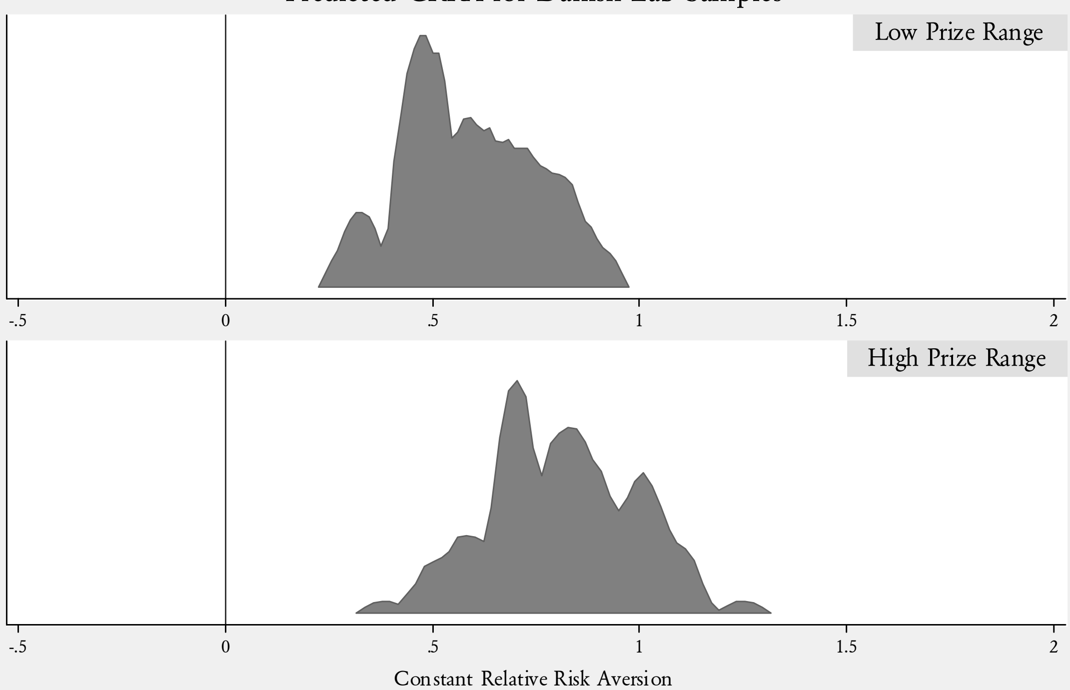


Figure 2: Effect of Varying the Range of Prizes

Predicted CRRA for Danish Lab Samples



References

Andersen, Steffen; Harrison, Glenn W.; Lau, Morten I., and Rutström, E. Elisabet, "Eliciting Risk and Time Preferences, *Working Paper 05-24*, Department of Economics, College of Business Administration, University of Central Florida, 2005.

Andersen, Steffen; Harrison, Glenn W.; Lau, Morten I., and Rutström, E. Elisabet, "Elicitation Using Multiple Price Lists," *Experimental Economics*, 9(4), December 2006, 383-405.

Camerer, Colin, and Lovallo, Dan, "Overconfidence and Excess Entry: An Experimental Approach," *American Economic Review*, 89(1), March 1999, 306-318.

Das, Mitali; Newey, Whitney K., and Vella, Francis, "Nonparametric Estimation of Sample Selection Models," *Review of Economic Studies*, 70, 2003, 33-58.

Diggle, P., and Kenward, M.G., "Informative Drop-out in Longitudinal Data Analysis," *Applied Statistics*, 43(1), 1994, 49-93.

Harrison, Glenn W.; Johnson, Eric; McInnes, Melayne M., and Rutström, E. Elisabet, "Temporal Stability of Estimates of Risk Aversion," *Applied Financial Economics Letters*, 2005a, 1(1), 31-35.

Harrison, Glenn W.; Johnson, Eric; McInnes, Melayne M., and Rutström, E. Elisabet, "Risk Aversion and Incentive Effects: Comment," *American Economic Review*, 95(3), June 2005b, 897-901.

Harrison, Glenn W.; Lau, Morten I; and Rutström, E. Elisabet, "Estimating Risk Attitudes in Denmark: A Field Experiment," *Scandinavian Journal of Economics*, 109, 2007 forthcoming.

Harrison, Glenn W.; Lau, Morten Igel; Rutström, E. Elisabet, and Sullivan, Melonie B., "Eliciting Risk and Time Preferences Using Field Experiments: Some Methodological Issues," in J. Carpenter, G.W. Harrison and J.A. List (eds.), *Field Experiments in Economics* (Greenwich, CT: JAI Press, Research in Experimental Economics, Volume 10, 2005).

Harrison, Glenn W.; Lau, Morten Igel, and Williams, Melonie B., "Estimating Individual Discount Rates for Denmark: A Field Experiment," *American Economic Review*, 92(5), December 2002, 1606-1617.

Harrison, Glenn W., and List, John A., "Field Experiments," *Journal of Economic Literature*, 42(4), December 2004, 1009-1055.

Heckman, James J., "The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependent Variables and a Simple Estimator for Such Models," *Annals of Economic and Social Measurement*, 5, 1976, 475-492.

Heckman, James J., "Sample Selection Bias as a Specification Error," *Econometrica*, 47, 1979, 153-161.

Heckman, James J., and Robb, Richard, "Alternative Methods for Evaluating the Impact of Interventions," in J. Heckman and B. Singer (eds.), *Longitudinal Analysis of Labor Market Data*

(New York: Cambridge University Press, 1985).

Heckman, James J., and Smith, Jeffrey A., "Assessing the Case for Social Experiments," *Journal of Economic Perspectives*, 9(2), Spring 1995, 85-110.

Holt, Charles A., and Laury, Susan K., "Risk Aversion and Incentive Effects," *American Economic Review*, 92(5), December 2002, 1644-1655.

Holt, Charles A., and Laury, Susan K., "Risk Aversion and Incentive Effects: New Data Without Order Effects," *American Economic Review*, 95(3), June 2005, 902-912.

Hotz, V. Joseph, "Designing an Evaluation of JTPA," in C. Manski and I. Garfinkel (eds.), *Evaluating Welfare and Training Programs* (Cambridge: Harvard University Press, 1992).

Kagel, John H.; Battalio, Raymond C., and Green, Leonard, *Economic Choice Theory: An Experimental Analysis of Animal Behavior* (New York: Cambridge University Press, 1995).

Kramer, Michael, and Shapiro, Stanley, "Scientific Challenges in the Application of Randomized Trials," *Journal of the American Medical Association*, 252, November 16, 1984, 2739-2745.

Lazear, Edward P.; Malmendier, Ulrike, and Weber, Roberto A., "Sorting in Experiments with Application to Social Preferences," *Unpublished Manuscript*, Graduate School of Business, Stanford University, October 2006.

Liang, K-Y., and Zeger, S. L., "Longitudinal Data Analysis Using Generalized Linear Models," *Biometrika*, 73, 1986, 13-22.

Palfrey, Thomas R., and Pevnitskaya, Svetlana, "Endogenous Entry and Self-Selection in Private Value Auctions: An Experimental Study," *Journal of Economic Behavior & Organization*, 64, 2007 forthcoming.

Philipson, Tomas, and Hedges, Larry V., "Subject Evaluation in Social Experiments," *Econometrica*, 66(2), March 1998, 381-408.

Rogers, W. H., "Regression standard errors in clustered samples," *Stata Technical Bulletin*, 13, 1993, 19-23.

Rutström, E. Elisabet, "Home-Grown Values and the Design of Incentive Compatible Auctions," *International Journal of Game Theory*, 27(3), 1998, 427-441.

StataCorp, *Stata Statistical Software: Release 8.0* (College Station, TX: Stata Corporation, 2003).

Vella, Francis, "Estimating Models with Sample Selection Bias: A Survey," *Journal of Human Resources*, 33(1), 1998, 127-169.

Williams, Rick L., "A Note on Robust Variance Estimation for Cluster-Correlated Data," *Biometrics*, 56, June 2000, 645-646.