A Survey Selection Correction using Nonrandom Followup with an Application to the Gender Entrepreneurship Gap

Clint Harris[*] Jon Eckhardt[†] Brent Goldfarb[‡]

April 26, 2024

Abstract

Selection into samples undermines efforts to describe populations and to estimate relationships between variables. We develop a simple method for correcting for sample selection that explains differences in survey responses between early and late respondents with correlation between potential responses and preference for survey response. Our method relies on researchers observing the number of data collection attempts prior to each individual's survey response rather than covariates that affect response rates without affecting potential responses. Applying our method to a survey of entrepreneurial aspirations among undergraduates at University of Wisconsin-Madison, we find suggestive evidence that the entrepreneurial aspiration rate is larger among survey respondents than the population, as well as the male-female gender gap in the entrepreneurial aspiration rate, which we estimate as 21 percentage points in the sample and 19 percentage points in the population. Our results suggest that the male-female gap in entrepreneurial aspirations arises prior to direct exposure to the labor market.

JEL Codes: C83, L26, J16

Keywords: Survey Nonresponse, Attrition, Sample Selection, Entrepreneurship, Gender

[*]University of Wisconsin-Madison, clint.harris@wisc.edu
[†]University of Wisconsin-Madison, jon.eckhardt@wisc.edu
[‡]University of Maryland, brentg@umd.edu

1 Introduction

Estimates of relationships between variables in a sample of a population are informative about the relationships between those variables in the population of interest when the sample is a random draw from the population (Fisher, 1922). To take advantage of this remarkable property, researchers commonly choose whether an individual in a population will be surveyed at random. Unfortunately, surveyed individuals choose whether they will respond to surveys nonrandomly. This nonrandom choice breaks the connection between between in-sample estimates and population estimands by introducing selection bias, which undermines inference about related research questions (Heckman, 1979).

This paper introduces a method for correcting for survey nonresponse by extrapolating from differences between early and late survey respondents in measured variables to differences between late and never responders in those variables. It has modest data demands, not requiring observation of baseline characteristics that predict survey nonresponse, unlike covariate adjustment based methods such as those described by Molina Millán and Macours (2017).[1] It also places minimal demands on the survey administrator, not requiring randomization of incentives or followup intensity, as in methods described by Dutz, Huitfeldt, Lacouture, Mogstad, Torgovitsky, and Van Dijk (2021) and DiNardo, Matsudaira, McCrary, and Sanbonmatsu (2021), respectively.[2] The present method may therefore be particularly attractive in applications using secondary data, where a researcher's estimation goals did not inform the data collection. Furthermore, it does not require financial or similar incentives, reducing costs (especially in large surveys) and avoiding the possibility of incentives crowding out intrinsic motivation for some individuals, causing them to reduce their response effort rather than increase it (Gneezy and Rustichini, 2000).

Our method shares all of the above advantages with that of Behaghel, Crépon, Gurgand, and Le Barbanchon (2015) for applications focused on treatment effect estimation. The modeling distinction between these methods is that ours requires assumptions (e.g. normality) on the joint distributions of unobserved determinants of the outcome of interest and survey response

[1] The method of this paper easily incorporates covariate adjustment if variables correlated with survey response are observed.

[2] Ex post randomization of which late responders' surveys to discard in a setting with multiple reminders for everyone would present a scenario in which the method of DiNardo, Matsudaira, McCrary, and Sanbonmatsu (2021) is directly applicable, with a slight tweak in the interpretation of who intensive followup compliers are. Our method relies on the same intuition without reductions in sample size from randomly discarding data (or never receiving it from followup compliers who are randomized into receiving low followup intensity).

aversion that are allowed to vary by treatment status, while theirs instead assumes that the distribution of survey response aversion for treated individuals is a monotonic transformation of the distribution for untreated individuals. Our alternative distributional assumptions lead to three key distinctions between results produced by our method and those produced by the method of Behagel et al. 2015. First, our method provides population level conditional expectation estimates, so it estimates average treatment effects for the population in treatment effect applications, rather than local average treatment effects specific to individuals with low survey response aversion. This distinction also extends our method's applicability to estimation of descriptive quantities such as population averages and comparisons between nonrandomly assigned groups (e.g. gender gaps), which may not have policy-relevant local averages. Second, our method produces point estimates of quantities of interest rather than bounds. Third, and finally, our method allows for heterogeneity between groups (such as by treatment status) in the correlation between outcomes and survey response aversion, which can drive differential selection bias between groups even in settings where they have similar survey response rates.[3]

We use our method to estimate the gender-gap in the prevalence of entrepreneurial aspirations among college students, using a survey of the undergraduate population of University of Wisconsin-Madison. A large body of work documents substantial differences in rates of entrepreneurship between men and women (Guzman and Kacperczyk, 2019), and suggests that women's underrepresentation in entrepreneurship may harm consumers if valuable products and services are overlooked by male entrepreneurs. Explanations for women's underrepresentation in entrepreneurship can be broadly categorized into nature vs. nurture, with nature-based explanations focusing on gender differences in relevant intrinsic preferences, such as for riskand competition, and with nurture-based explanations focusing on gender differences in external factors, such as labor market discrimination, differences in social norms surrounding childcare, and premarket disparities in education or encouragement relevant to entrepreneurship. Our nonresponse-corrected estimate of the entrepreneurial aspirations gender gap is large (19 percentage points) and of similar magnitude to the uncorrected estimate (21 percentage points), suggesting that premarket factors are major determinants of the entrepreneurship gender gap.

[3]Our method is robust to settings in which a treatment with no effect on outcomes increases (decreases) survey responses for high outcome (low outcome) individuals, which would likely yield positive effect estimates without correction or with a correction that relied on monotonic relationships between survey responses and outcomes between groups.

2 Survey Nonresponse Correction with Multiple Reminders

We consider a randomly surveyed sample of N individuals indexed by i who receive T requests to fill out a survey, with survey requests indexed by t. We denote individual i's potentially unobserved outcome of interest as Y_i and the binary indicator for them responding to request t or a prior request as S_{it}, such that $S_{iT} = 1$ if they ever respond and $S_{iT} = 0$ if they never respond. We are interested in estimation of quantities such as the unconditional expectation of the outcome in the population, $\mathbb{E}[Y_i]$, conditional expectations of the outcome for various groups defined by values of observed covariates such as $\mathbb{E}[Y_i|W_i = w]$, and differences between groups in their expected outcomes such as $\mathbb{E}[Y_i|W_i = w] - \mathbb{E}[Y_i|W_i = w']$, where $W_i \subseteq X_i$ and X_i is the set of covariates observed for surveyed individuals.[4]

2.1 Problem

In settings with survey nonresponse, the conditional sample mean among individuals for whom $W_i = w$ is a consistent estimate of $\mathbb{E}[Y_i|W_i = w, S_{iT} = 1]$, which may differ from $\mathbb{E}[Y_i|W_i = w]$ as described by Heckman (1979). The expected nonresponse bias for group $W_i = w$ is given by

$$B_w = \mathbb{E}[Y_i|W_i = w, S_{iT} = 1] - \mathbb{E}[Y_i|W_i = w]. \tag{1}$$

The difference in sample means between groups defined by $W_i = w$ and $W_i = w'$ provides an estimate of

$$\begin{aligned} &\mathbb{E}[Y_i|W_i = w, S_{iT} = 1] - \mathbb{E}[Y_i|W_i = w', S_{iT} = 1] \\ =&\mathbb{E}[Y_i|W_i = w] - \mathbb{E}[Y_i|W_i = w'] + B_w - B_{w'}, \end{aligned} \tag{2}$$

which is only equal to the population difference between groups if their survey nonresponse bias terms are equal. In settings with 100% survey response rates (or random nonresponse), each of these bias terms is equal to zero. It follows that in such cases we can consistently estimate $\mathbb{E}[Y_i|W_i = w]$ for any $W_i \subseteq X_i$ and w in the support of W_i by calculating the relevant conditional sample mean, and we can compare or average these quantities among groups as desired.

[4] The comparison $\mathbb{E}[Y_i|W_i = w] - \mathbb{E}[Y_i|W_i = w']$ gives the average treatment effect of $W_i = w$ versus $W_i = w'$ in cases in which W_i is randomized, with similar quantities being relevant to other research designs focused on measuring treatment effects. The application in this paper investigating the gender gap in entrepreneurial aspirations estimates such a quantity, following procedures that would consistently estimate the treatment effect of gender if it were randomly assigned.

2.2 Solution

It is hard to justify a priori assumptions about nonresponse bias magnitudes or random nonresponse, so we propose a method for extrapolating from observed values of Y_i for individuals who respond to the survey to the rest of the population. In our application, outcomes of interest are binary choices made by agents, so we assume that

$$Y_i = \begin{cases} 1 \text{ if } X_i\beta + \epsilon_i \geq 0, \\ 0 \text{ otherwise,} \end{cases} \tag{3}$$

wherein X_i is a $1 \times K$ vector containing a constant and observed variables that predict outcomes, β is an $K \times 1$ vector of marginal effects of variables in X_i, and ϵ_i is a mean zero residual.[5] Importantly, the outcome process in (3) contains no t subscripts, so effects of notifications (or time between them) on the outcome or its measurement at various points in time are assumed to be zero. We similarly assume that individuals make a binary choice regarding responding to survey request t or a prior request according to the rule

$$S_{it} = \begin{cases} 1 \text{ if } Z_i\alpha_t + u_i \geq 0, \\ 0 \text{ otherwise,} \end{cases} \tag{4}$$

wherein Z_i is a $1 \times L$ vector containing a constant and observed variables that predict outcomes, α_t is an $L \times 1$ vector of marginal effects of variables in Z_i, and u_i is a mean zero residual.[6]

We follow Heckman (1979) and Van de Ven and Van Praag (1981), who extended Heckman's selection correction method to the case with a binary outcome, in assuming that unobserved residuals of binary outcome equations are normally distributed, in our case as

$$\begin{bmatrix} \epsilon_i \\ u_i \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right). \tag{5}$$

We discuss the implications of these assumptions relative to some alternatives in Section 2.3, and we describe tests of these assumption that can be performed in applications with more than

[5]The case in which Y_i is continuous will be described in a future Appendix.

[6]A similar model would treat the response decision as ordered and would assume individual i responds after notification t and before notification $t+1$ if $\mu^t \leq Z_i\alpha + u_i < \mu^{t+1}$ with normalizations $\mu^1 = 0$ and $\mu^{T+1} = \text{inf}$. The model in (4) is attractive because it allows for individuals with different values of Z_i to differ in their intensities of response to reminders, essentially using $\mu^t(Z_i) = Z_i\mu^t$ in the ordered model and defining $\alpha_t = \alpha - \mu^t$.

one reminder in Section 2.4. Under this assumption, the log-likelihood for the sample is

$$\begin{aligned}
&\mathcal{L}\left(\beta, \alpha_1, ..., \alpha_T, \rho | X_i, Z_i\right) = \\
&\sum_{t=1}^{T}\sum_{i=1}^{N} S_{it}(Y_i \ln(\Phi_2(X_i\beta, Z_i\alpha_t, \rho)) + (1-Y_i)\ln(\Phi_2(-X_i\beta, Z_i\alpha_t, -\rho))) \\
&+(1-S_{it})\ln(1-\Phi(Z_i\alpha_t)),
\end{aligned} \tag{6}$$

wherein $\Phi(\cdot)$ is the standard normal CDF and $\Phi_2(\cdot)$ is the standard bivariate normal CDF. The values $(\hat{\beta}, \hat{\alpha}_1, ..., \hat{\alpha}_T, \hat{\rho})$ that maximize this likelihood are estimates of their corresponding population parameters, and the estimated conditional probability of Y_i given X_i is $\Phi(X_i\hat{\beta})$.

It is helpful to think through the estimator's performance in a few extreme cases to build intuition. First, in the case where $t = 1$ and $X_i = Z_i$, maximizing the likelihood subject to the constraint $\rho = 0$ provides the same estimates of β as a probit regression of Y_i on X_i, which is valid if survey nonresponse is random conditional on X_i. Without an unconstrained ρ parameter, the estimator reduces to the probit version of Heckman's sample selection correction method, which is reliable for $T = 1$ in cases where there is at least one variable in Z_i that is not also in X_i. This motivates the use of randomized survey response incentives or followup intensity as advocated by Dutz, Huitfeldt, Lacouture, Mogstad, Torgovitsky, and Van Dijk (2021) and DiNardo, Matsudaira, McCrary, and Sanbonmatsu (2021). In cases where $t > 1$ and $X_i = Z_i$, the model is identified by differences in average values of Y_i for individuals who respond to the survey at different times. This is equivalent to Heckman's correction with $X_i \subset Z_i$ under an equivalent model that defines $Z_i = [X_i, X_iS_{i2}, ...X_iS_{iT}]$, with an $1 \times KT$ coefficient vector α present in survey response equations for all survey notifications.

2.3 Assumption Violation Implications

It is useful to think through implications of violations of model assumptions. First, if the outcome (or nonclassical error in its measurement) changes in response to the passage of time or in response to survey reminders, at odds with the time invariance assumption implied by (3), then the estimated model's attribution of differences between early and late responders to survey response preferences will be inaccurate. Concerns relating to the passage of time can be eliminated if survey responses are restricted to time invariant quantities, which survey administrators can encourage by careful wording of questions.[7] Direct effects of reminders on time-invariant outcomes seem unlikely, but effects on their measurement could occur if reminders affect survey effort. For instance, sending a large number of reminders in quick succession could annoy respondents and cause them to provide responses that are uncorrelated or even negatively correlated with true values.

Second, if the distributional assumption on ϵ_i and u_i in (5) is inaccurate, the out of sample extrapolation will suffer. In cases with a single reminder, estimates of conditional in-sample means given by $\Phi(X_i\hat{\beta}+\hat{\rho}\lambda(Z_i\hat{\alpha}))$ (where $\lambda(Z_i\alpha)$ is the inverse-Mills ratio evaluated at $Z_i\alpha$) for early and late responders will be equal to sample means for both subgroups (the constant and the parameter ρ are chosen to ensure this), but out-of-sample fit may be poor. In cases with more than one reminder, the in-sample fit could suffer as well, which motivates the test of this assumption described in section 2.4.

2.4 Testing

In settings with more than one reminder, the model is overidentified. Assuming that the passage of time and the receipt of notifications are excludeable from the outcome equation, the bivariate normality assumption on model residuals given in (5) can be assessed by estimating the log-Likelihood

$$\begin{aligned}\mathcal{L}\left(\beta,\alpha_1,...,\alpha_T,\rho_2,...,\rho_T|X_i,Z_i\right) &= \sum_{t=1}^{T}\sum_{i=1}^{N}(1-S_{it})\ln(1-\Phi(Z_i\alpha_t))\\ +\sum_{t=1}^{2}\sum_{i=1}^{N}S_{it}&\left(Y_i\ln(\Phi_2(X_i\beta,Z_i\alpha_t,\rho_2))+(1-Y_i)\ln(\Phi_2(-X_i\beta,Z_i\alpha_t,-\rho_2))\right)\\ +\sum_{t=3}^{T}\sum_{i=1}^{N}S_{it}&\left(Y_i\ln(\Phi_2(X_i\beta,Z_i\alpha_t,\rho_t))+(1-Y_i)\ln(\Phi_2(-X_i\beta,Z_i\alpha_t,-\rho_t))\right),\end{aligned} \tag{7}$$

and testing the hypothesis $\hat{\rho}_2=\hat{\rho}_t$ for $t=3,...,T$. This test essentially compares the correlation parameter, $\hat{\rho}_2$, that rationalizes the differences in outcomes between individuals who respond to the first or second notification to the correlation parameters that rationalize outcomes for those who respond to notification t or before, for all $t>2$, given the estimate of β from the first two

[7] To use a salient example, we would expect our method to perform poorly if used during a pandemic on a survey asking individuals if they currently have the illness because average values of this quantity would change over time. We would expect it to perform better if used on a similar survey asking individuals if they had the illness as of a specified date (such as the date the survey was sent out).

notifications.[8] The equality of these correlation parameters is necessary, but not sufficient, to determine that the true distribution of errors is bivariate normal.

This test is unduly strict for some applications. It will reject the null of bivariate normality in cases in which the correlation parameter that rationalizes outcomes at $t = 1$ and $t = 2$ fails to rationalize outcomes at some larger values of t, even if it does accurately rationalize later outcomes *on average*. If the model's prediction errors for outcomes from estimates only using the first two notifications are independent of t in and out-of-sample, model estimates of population expectations will be consistent, with reliability increasing in the number of reminders (such that the $\hat{\rho}$ of best fit averages across the prediction errors rather than fitting them).[9] A more lenient test can be implemented by maximizing

$$\begin{aligned}\mathcal{L}\left(\beta, \alpha_1, ..., \alpha_T, \rho_2, ..., \rho_T | X_i, Z_i\right) &= \sum_{t=1}^{T}\sum_{i=1}^{N}(1 - S_{it})\ln(1 - \Phi(Z_i\alpha_t)) \\ + \sum_{t=1}^{2}\sum_{i=1}^{N} S_{it}\left(Y_i \ln(\Phi_2(X_i\beta, Z_i\alpha_t, \rho_2)) + (1 - Y_i)\ln(\Phi_2(-X_i\beta, Z_i\alpha_t, -\rho_2))\right) \\ + \sum_{i=1}^{N} S_{iT}\left(Y_i \ln(\Phi_2(X_i\beta, Z_i\alpha_T, \rho_T)) + (1 - Y_i)\ln(\Phi_2(-X_i\beta, Z_i\alpha_T, -\rho_T))\right),\end{aligned} \tag{8}$$

and testing the hypothesis $\hat{\rho}_2 = \hat{\rho}_T$. This test checks whether the estimated conditional expectations of Y_i and correlation parameter obtained using only the first two notifications also rationalizes average observed outcomes among all individuals who respond to the survey, without penalizing prediction error for distinct response groups.

3 Application: Gender Entrepreneurship Gap

We use our method to estimate the gender gap in entrepreneurial aspirations among college students. There is a large gender gap in entrepreneurship between working age adults (Aldrich, 2005), driven at least in part by a gender gap in early stage funding for new ventures (Canning, Haque, and Wang, 2012; Greene, Hart, Gatewood, Brush, and Carter, 2003). If there is no such gap in intentions among individuals prior to entering the labor market, we can conclude that

[8] For instance, when $X_i = Z_i = 1$, α_t is exactly identified by the survey response rate up to reminder t, β is identified as the predicted population expectation of Y_i using only the first two notifications, and ρ_t for $t > 2$ rationalizes the within-sample average of Y_i up to notification t taking $\hat{\beta}$ as given from the first two notifications.

[9] This strictness is similar to that of pre-trend tests used in difference-in-differences applications that check whether *all* pre-treatment deviations from parallel trends are equal to zero. In such cases estimates of policy effects that average over a large number of post-treatment time periods are valid if parallel trends violations are independent of time.

the gap is due to market forces and not gender differences in other factors such as personal preferences relevant to entrepreneurship or premarket discrimination. We contribute to this literature by estimating the entrepreneurial aspiration gender gap using the method described in the previous section to address selection bias.

3.1 Data

We use a survey of the undergraduate population of University of Wisconsin - Madison that was implemented every Fall from 2015 to 2022, as well as in Spring 2020 and 2021. We match the survey data to administrative data on students, which contains information on individuals regardless of whether they responded to the survey. This survey asked students whether they intended to pursue a career in entrepreneurship. All responses were given as "Yes", "No", or "I Don't Know", where we code responses as binary based on whether the answer given was "Yes". Summary variables for those surveyed broken down by if and when they responded are provided in Table 1.

Table 1: Summary Statistics

	Early Respondents		Late Respondents		Nonrespondents	
	Men	Women	Men	Women	Men	Women
	(1)	(2)	(3)	(4)	(5)	(6)
Intention	0.374	0.167	0.363	0.179	.	.
Business Major	0.128	0.093	0.130	0.098	0.125	0.086
STEM Major	0.455	0.276	0.455	0.269	0.432	0.261
ACT Math	30.220	28.025	30.252	27.912	29.793	27.520
ACT Verbal	55.718	56.478	54.747	55.667	54.239	54.798
Racial Minority	0.242	0.227	0.287	0.264	0.302	0.281
International	0.098	0.063	0.108	0.070	0.103	0.071
First-Gen College	0.142	0.177	0.144	0.180	0.154	0.193
In-State	0.617	0.649	0.576	0.609	0.546	0.562
Year 2 Student	0.220	0.219	0.229	0.230	0.248	0.258
Year 3 Student	0.176	0.183	0.184	0.191	0.236	0.245
Year 4 Student	0.168	0.181	0.154	0.164	0.204	0.211
Year 5+ Student	0.059	0.037	0.055	0.037	0.069	0.042
2016, Fall	0.128	0.135	0.092	0.101	0.103	0.096
2017, Fall	0.152	0.167	0.074	0.083	0.102	0.096
2018, Fall	0.076	0.081	0.112	0.130	0.100	0.100
2019, Fall	0.082	0.088	0.079	0.085	0.105	0.107
2020, Spring	0.065	0.073	0.112	0.114	0.095	0.098
2020, Fall	0.104	0.091	0.161	0.161	0.099	0.101
2021, Spring	0.068	0.065	0.083	0.079	0.100	0.103
2021, Fall	0.085	0.072	0.143	0.119	0.099	0.103
2022, Fall	0.044	0.037	0.100	0.079	0.100	0.102
Observations	9563	12407	8703	10152	106582	114979

Notes: Variable means for men and women by responder group. Early Respondents are individuals who respond prior to any reminders, while Late Respondents are individuals who respond any time after the first reminder.

3.2 Estimated Quantities

We begin by estimating the rate of entrepreneurial intention among men and women separately in the highest response rate term (Fall, 2020), without using the predictor variables listed in Table 1. This parsimonious specification is pedagogically attractive because it restricts attention to our methodological contribution, while also clarifying the close connection of our method to a Heckman correction, especially in the specific case of randomized followup intensity as proposed by DiNardo, Matsudaira, McCrary, and Sanbonmatsu (2021). Specifically, we maximize the log-Likelihood in (6) and calculate $\Phi(\hat{\beta}_0)$ separately for men and women, where $\hat{\beta}_0$ is the coefficient on the constant. We also calculate $\Phi(\hat{\beta}_0 + \hat{\rho}\Phi^{-1}(1 - e_i))$ for the range of values of $e_i \in [0, 1]$ separately for men and women, in order to trace out predicted entrepreneurial aspirations for individuals across percentiles of survey response aversion. Importantly, we always estimate gender-specific residual correlation (ρ) parameters.[10]

We also estimate the average gender gap, and we perform an Oaxaca-Blinder decomposition to identify variables that are relevant to the gender gap (Oaxaca, 1973; Blinder, 1973). This method considers, for each covariate in X_i that determines outcomes, the extent to which the gender gap would be reduced if women were like men with respect to that single variable, and how much it would be reduced if women were like men with respect to how much that variable determines outcomes (given by the element of β corresponding to the variable in a model where X_i).

The counterfactual in which women are "like men" with respect to a given variable or its coefficient can be interpreted multiple ways. Specifically, we define F_i as a binary indicator for female gender, and we estimate versions of outcome process (3) of the form

$$Y_i = X_i\beta + (F_i \times X_i)\delta + \epsilon_i \tag{9}$$

where $\times$ defines the dot product such that $F_i \times X_i$ gives the interaction of the female indicator and other variables, with X_i here defined differently from above as the variables other than gender that are assumed to determine the outcome. For a single variable $W_i \subseteq X_i$, we estimate

[10] Allowing group specific residual correlations is potentially important in applications involving treatment effect estimation. For instance, in a randomized controlled trial with post-experiment followup attrition, allowing ρ to be different for treated and untreated individuals can correct for the possibility of differential selection bias, which is arguably the main threat to identification in such settings.

the contribution to the average gender gap of that variable as

$$\hat{\Delta}_W = \sum_{i=1}^{N} \frac{\Phi(X_i\hat{\beta} + (F_i \times X_i)\hat{\delta} + F_i \times (\bar{W}^M - \bar{W}^F)(\hat{\beta}_W + \hat{\delta}_W)) - \Phi(X_i\hat{\beta} + (F_i \times X_i)\hat{\delta})}{N}, \quad (10)$$

wherein $\bar{W}^M$ is the sample mean of W_i for men, $\bar{W}^F$ is the sample mean of W_i for women, $\hat{\beta}_W$ is the coefficient on W_i and $\hat{\delta}_W$ is the coefficient on the interaction between W_i and F_i.[11] We similarly estimate the contribution to the average gender gap of that variable's differential effect on Y_i between men and women as

$$\hat{\Delta}_{\beta_W} = \sum_{i=1}^{N} \frac{\Phi(X_i\hat{\beta} + (F_i \times X_i)\hat{\delta} - (F_i \times W_i^F)\hat{\delta}_W) - \Phi(X_i\hat{\beta} + (F_i \times X_i)\hat{\delta})}{N}. \quad (11)$$

3.3 Results

We begin by presenting estimates of average rates of entrepreneurial intention among men and women for Fall, 2020 without using the predictor variables listed in Table 1. We present these results graphically in Figures 1 for men and Figure 2 for women. The top panel of both figures shows means within each response group, with individuals responding to the third or later reminder pooled together for the graph, but not for the estimation. The top panel also includes uncorrected and corrected extrapolations using within-sample means (equivalent to results from an in-sample probit with only a constant) and our method, respectively. The bottom panel of each figure adds extrapolated estimates from the overidentification test from maximizing the likelihood in (7), with overlapping confidence intervals for all extrapolations showing that, within this one year, we are unable to reject the uncorrected method null of no selection bias or the corrected method null of bivariate normality on residuals for both men and women.

We present estimates of the average entrepreneurial aspiration rate for men and women for the entire population, using all years of our survey, in Table 2. This table also includes the estimated gender gaps in aspirations from the uncorrected and corrected method, along with a decomposition of the contributions of variables and their coefficients to the gender gap, calculated via equations (10) and (11), respectively.

The panel denoted X shows mean values for each variable in X_i for both men and women

[11] This is simpler than the commonly used extension of the Oaxaca-Blinder Decomposition developed by Fairlie (2005) for nonlinear models. Our method provides predictions (under conditional independence and homogeneous effects assumptions on W_i) for the gap in Y_i under counterfactual policies that change W_i the same amount for all women. Fairlie's method provides predictions (under the same assumptions) for the gap in Y_i under more nuanced counterfactual policies that shift W_i different amounts for different women to match men's marginal distribution of W_i.

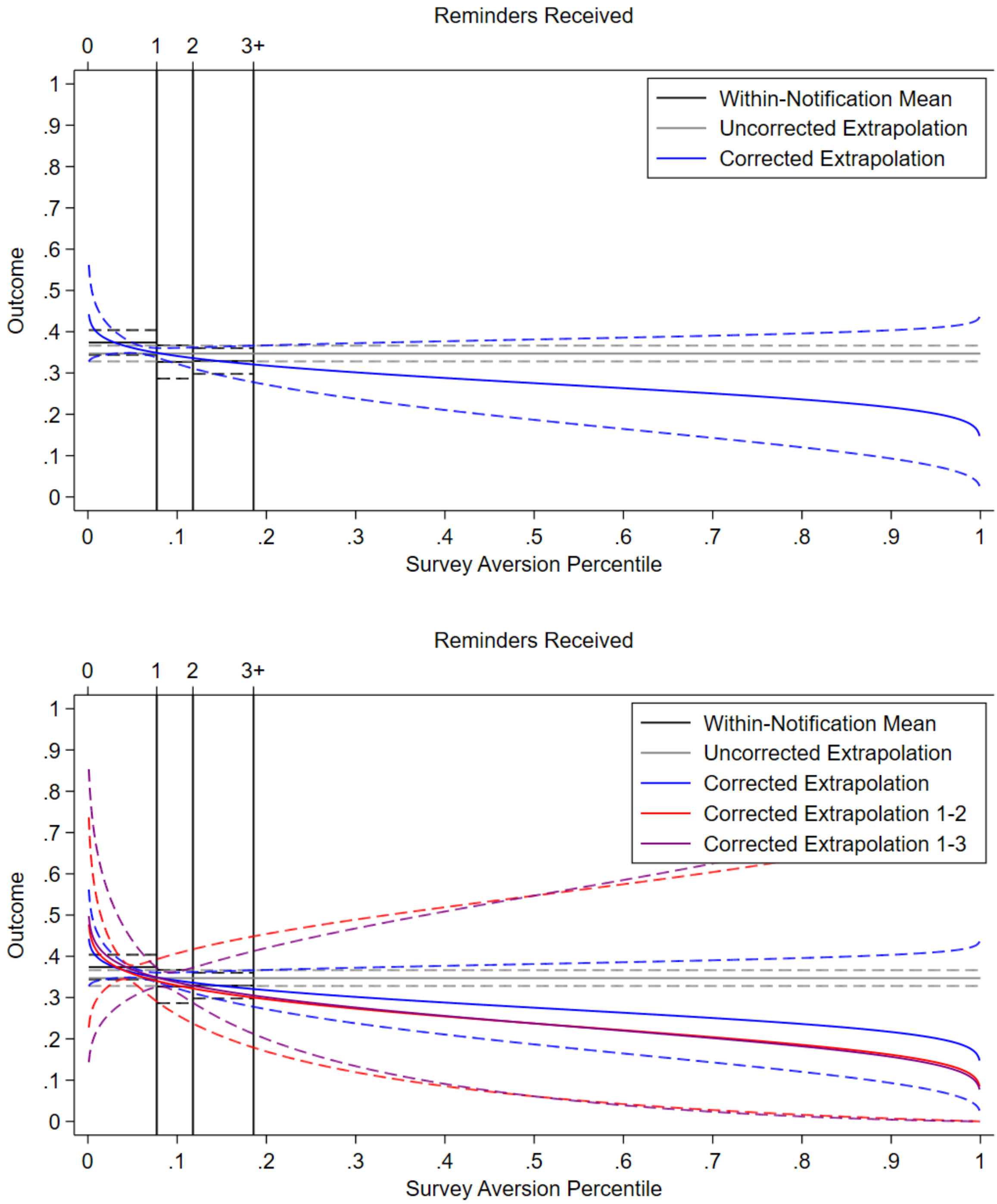

Figure 1: Entrepreneurial Intentions for Men

Notes: Estimated average entrepreneurial aspiration rates. Uncorrected estimates are calculated as the sample average among all survey respondents. Corrected estimates are calculated as $\Phi(\hat{\beta}_0 + \hat{\rho}\Phi^{-1}(1 - e_i))$ for the range of values of survey aversion percentiles, $e_i \in [0, 1]$.. Corrected extrapolations 1-2 and 1-3 provide estimates of $\Phi(\hat{\beta}_0 + \hat{\rho}_2\Phi^{-1}(1 - e_i))$ and $\Phi(\hat{\beta}_0 + \hat{\rho}_3\Phi^{-1}(1 - e_i))$, respectively, estimated by maximizing the likelihood in (7).

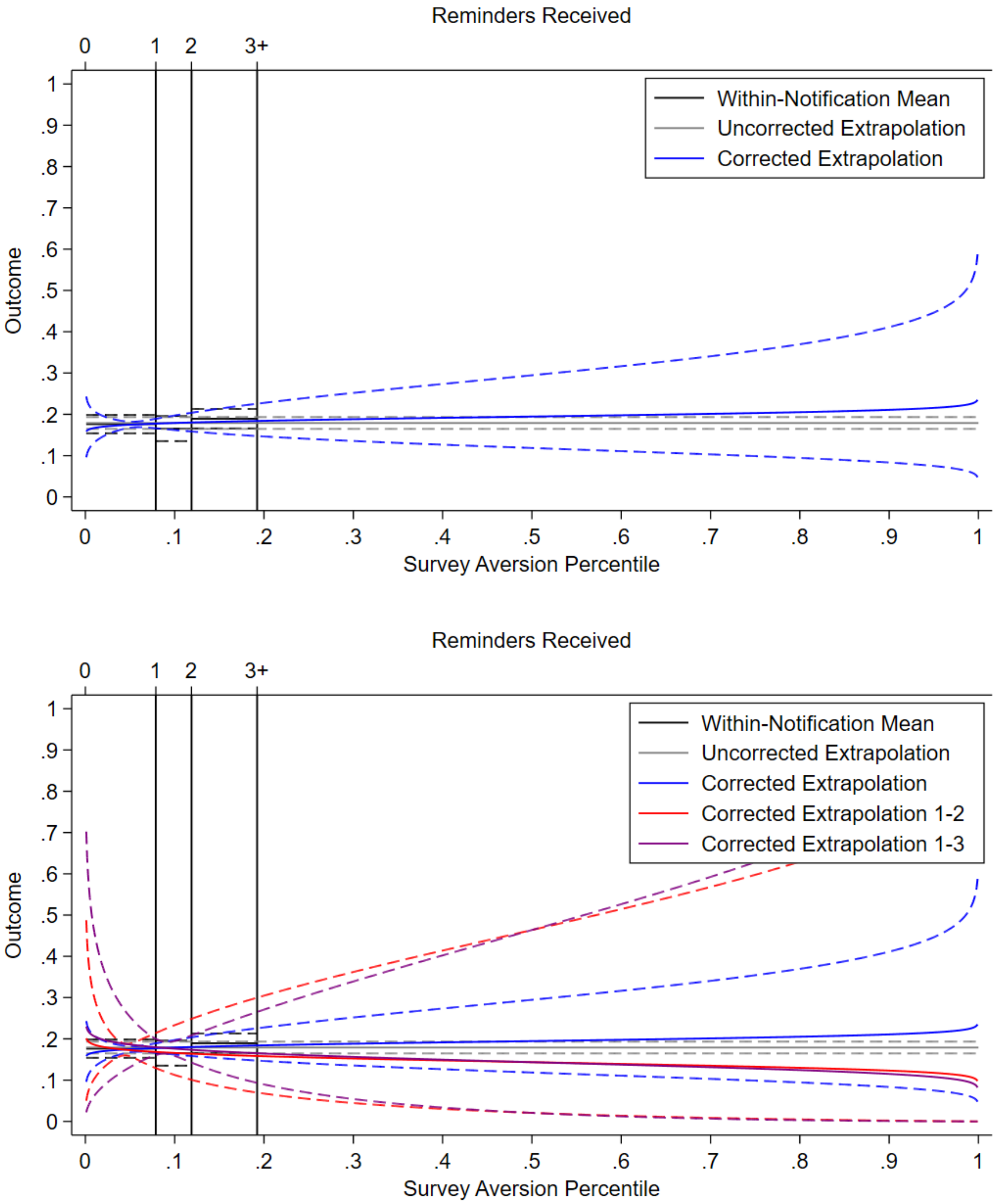

Figure 2: Entrepreneurial Intentions for Women

Notes: Estimated average entrepreneurial aspiration rates. Uncorrected estimates are calculated as the sample average among all survey respondents. Corrected estimates are calculated as $\Phi(\hat{\beta}_0 + \hat{\rho}\Phi^{-1}(1 - e_i))$ for the range of values of survey aversion percentiles, $e_i \in [0, 1]$. Corrected extrapolations 1-2 and 1-3 provide estimates of $\Phi(\hat{\beta}_0 + \hat{\rho}_2\Phi^{-1}(1 - e_i))$ and $\Phi(\hat{\beta}_0 + \hat{\rho}_3\Phi^{-1}(1 - e_i))$, respectively, estimated by maximizing the likelihood in (7).

who responded to the survey in columns (1) and (2). Columns (4) and (5) provide the sample means of each variable for men and women for all individuals in our data regardless of whether they responded to the survey. Columns (3) and (6) report estimated contributions of each variable in X_i to the gender gap calculated as described in equation (10) using values of X_i for all individuals in our data, with the uncorrected estimates using $\hat{\beta}$ from a within-survey probit and the corrected estimates coming from maximizing the log-Likelihood in (6).

The panel denoted β shows the gender-specific estimates of β from a probit in columns (1) and (2), and the gender-specific estimates of β from our new method in columns (4) and (5). The estimated gap contributions of each element of β, calculated via (11) on the full sample of individuals surveyed, are given in column (3) for the uncorrected method and (6) for the corrected method. We also report the effect of shifting all variables in X for women to match their male means, as well as shifting all values of β for women to match those of men. The "unexplained gap" is the additional difference in estimated average aspirations among men and women that is not explained purely by X_i or β, which is partially driven by the interaction of the gender gap in X_i and the gender gap in β, and is also partially driven by nonlinear effects of non-mean differences in the distribution of X_i between men and women.

In addition to the contributions of gender gaps in each variable and its coefficient to the gap in entrepreneurial aspirations, we also report gender-specific estimates of the correlation parameter, $\hat{\rho}$, in columns (4) and (5), with the constrained values of zero implied by the uncorrected method shown in columns (1) and (2). With multiple reminders, ρ can be estimated, so the p-value on $\hat{\rho}$ in the corrected model is a test of the null hypothesis of no correlation in unobservables, as assumed by the uncorrected method. This "overidentification-test" p-value is reported for men and women in columns (1) and (2), with the overidentification-test p-value from the test described in Section 2.4 shown in columns (4) and (5). We fail to reject the null $\rho = 0$ for both men and women at conventional significance levels with p-values of 0.129 for both groups. We reject the assumption of joint normality in (5) for men with a p-value of 0.040, while failing to reject this assumption for women, with a p-value of 0.376.

Both methods estimate large gaps between men and women in entrepreneurial aspirations, with the uncorrected and corrected gap estimates given by 0.207 and 0.194, respectively. We find marginally significant evidence of positive selection bias for both men and women with p-values of 0.103 and 0.090, respectively, which suggests that individuals who aspire to become entrepreneurs may be more likely to respond to a survey about entrepreneurship. The difference between the uncorrected and corrected gap estimates of 0.013 (0.207-0.194) is also just shy of conventional significance thresholds, with a p-value of 0.103, suggesting that the sample selection bias might be larger for men than for women. This is consistent with a survey response decision that is driven both by interest in entrepreneurship (higher for men, lower for women) and conscientiousness (lower for men, higher for women).

We do not find any particularly important predictors of entrepreneurial aspirations among our predictor variables, with the total estimated contribution of X_i at 1.1 percentage points for both the uncorrected and corrected models. The total uncorrected and corrected effects of β (0.189 and 0.179) on the gender gap are overwhelmingly driven by the contributions of the constants, which are 0.154 and 0.172 for the uncorrected and corrected methods, respectively.

Overall, we find evidence of large gender gaps in entrepreneurial aspirations among college students using a survey of undergraduate students at the university of Wisconsin - Madison. The average gap among the survey respondents is slightly smaller than the estimates provided by controlling for observed determinants of survey response, with both of these slightly larger than the gap estimates obtained using our method to control for otherwise unobserved determinants of response using the timing of responses. We note that we reject the bivariate normality assumption on model residuals for men, so we take these results with a grain of salt. We plan to consider extensions of our method in a future version of this paper that considers alternative distributional assumptions when bivariate normality is rejected.

4 Conclusion

We develop a method for estimating population level relationships between variables using survey data that corrects for selection into response that can be used in settings with multiple reminders. The core insight is to estimate differences in survey variables between early and late responders, and to extrapolate from those differences to the rest of the population of interest, for whom survey data does not exist. The method nests a standard Heckman correction as a special case in which there is a single survey notification. This means it corrects for both observed and unobserved determinants of selection by permitting, but not relying on, the use of covariates that affect both the variable surveyed and the survey response decision as well as those that only affect the survey response decision, such as randomized response incentives.

Our method is particularly noteworthy because of its low costs in multiple dimensions.

Table 2: Oaxaca-Blinder Decomposition of the Entrepreneurial Aspirations Gender Gap

	Uncorrected			Corrected		
	Male	Female	Gap Portion	Male	Female	Gap Portion
Variables	(1)	(2)	(3)	(4)	(5)	(6)
Intention	0.389	0.182	0.207	0.352	0.158	0.194
	(0.005)	(0.004)	(0.006)	(0.024)	(0.015)	(0.028)
X						
Business Major	0.128	0.096	0.004	0.121	0.085	0.004
	(0.002)	(0.002)	(0.000)	(0.001)	(0.001)	(0.000)
STEM Major	0.433	0.256	-0.002	0.417	0.249	-0.002
	(0.004)	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)
GPA	3.408	3.438	0.005	3.288	3.392	0.004
	(0.004)	(0.003)	(0.001)	(0.002)	(0.001)	(0.001)
ACT Math	30.156	27.842	-0.005	29.826	27.526	-0.003
	(0.028)	(0.026)	(0.002)	(0.011)	(0.011)	(0.002)
ACT Verbal	54.864	55.810	0.002	54.026	54.819	0.002
	(0.086)	(0.072)	(0.000)	(0.033)	(0.030)	(0.000)
Racial Minority	0.268	0.243	0.001	0.301	0.271	0.001
	(0.003)	(0.003)	(0.000)	(0.001)	(0.001)	(0.000)
International	0.099	0.064	0.004	0.101	0.068	0.004
	(0.002)	(0.002)	(0.000)	(0.001)	(0.001)	(0.001)
All X			0.011			0.011
			(0.003)			(0.003)
β						
Business Major	0.442	0.497	-0.002	0.452	0.506	-0.001
	(0.040)	(0.042)	(0.002)	(0.042)	(0.045)	(0.002)
STEM Major	-0.033	-0.057	0.001	-0.040	-0.047	0.000
	(0.028)	(0.032)	(0.003)	(0.031)	(0.035)	(0.003)
GPA	-0.173	-0.183	0.008	-0.144	-0.147	0.002
	(0.028)	(0.031)	(0.036)	(0.032)	(0.034)	(0.036)
ACT Math	-0.005	-0.009	0.028	-0.005	-0.006	0.000
	(0.004)	(0.004)	(0.044)	(0.004)	(0.005)	(0.040)
ACT Verbal	-0.007	-0.008	0.018	-0.007	-0.009	0.024
	(0.001)	(0.001)	(0.027)	(0.001)	(0.001)	(0.027)
Racial Minority	0.090	0.149	-0.005	0.071	0.150	-0.006
	(0.036)	(0.035)	(0.004)	(0.039)	(0.038)	(0.004)
International	0.290	0.543	-0.006	0.344	0.505	-0.004
	(0.055)	(0.060)	(0.002)	(0.061)	(0.065)	(0.002)
Constant	0.860	0.350	0.154	0.702	0.110	0.172
	(0.143)	(0.143)	(0.070)	(0.182)	(0.178)	(0.086)
All β			0.189			0.179
			(0.007)			(0.028)
All Unexplained Gaps			0.006			0.004
			(0.005)			(0.005)
Auxiliary Parameters						
ρ	0	0	0	0.061	0.064	0
	(.)	(.)	(.)	(0.040)	(0.042)	(.)
Log-Likelihood	-299757.3					
p-val overidentification	0.129	0.129		0.040	0.376	
p-val, Difference from Corrected	0.103	0.090	0.103			
Observations (responded)	18227	22523	40750	18227	22523	40750
Observations (surveyed)	124830	137524	262354	124830	137524	262354

Notes: X estimates in columns 1-2 are means among individuals who responded to the survey, with columns 4-5 giving means among all individuals surveyed. Columns 3 and 6 give uncorrected and selection-corrected estimated contributions of each variable to the gap in entrepreneurial intentions from equation (10). β estimates give the estimated coefficients on variables from a probit in columns 1-2, and from the selection correction in columns 4-5. Columns 3 and 6 give uncorrected and selection-corrected estimated contributions of each value in β to the gap in entrepreneurial intentions from equation (11). p-values from the overidentification test estimated via (7) are reported for men and women along with p-values on the difference between corrected and uncorrected estimates of the average entrepreneurial intention rate.

First, it does not require observation of any variables other than those gathered during survey administration responses for any surveyed individuals, setting it apart from methods that adjust for observable determinants of survey response. Second, it does not require randomized survey response incentives, which minimizes both administrative complexity and financial costs, as well as making it ideal for secondary data from surveys that were not administered with the goal of eliminating selection bias. Finally, it relies on a straightforward alteration of an existing method that is familiar to many applied researchers, implemented in statistical software as a loop over Heckman corrections for each reminder.

Following Heckman (1979) and Van de Ven and Van Praag (1981), we model all binary variables as arising from latent variable processes with normally distributed residuals. The core of our method works with any assumed bivariate distribution for residuals in the survey response decision and the outcome of interest. The range of alternative methods includes methods that avoid distributional assumptions at the cost of obtaining bounds rather than point estimates on quantities of interest, such as that of Lee (2005) or Behaghel, Crépon, Gurgand, and Le Barbanchon (2015). It also includes methods that allow richer structure on the survey selection choice, such as the method of Dutz, Huitfeldt, Lacouture, Mogstad, Torgovitsky, and Van Dijk (2021), at the cost of requiring variables that affect the survey response decision but not the outcome of interest in addition to survey reminders. We leave a survey of the performance of alternative methods and distributional assumptions across a variety of applications to future work, noting that the best method for a particular application will be the one that leverages subject matter expertise to obtain good in-sample fit and out-of-sample fit subject to constraints imposed by data availability.

References

Aldrich, H. (2005): "Entrepreneurship," *Research in the Sociology of Work*, 15, 3–31.

Behaghel, L., B. Crépon, M. Gurgand, and T. Le Barbanchon (2015): "Please call again: Correcting nonresponse bias in treatment effect models," *Review of Economics and Statistics*, 97(5), 1070–1080.

Blinder, A. S. (1973): "Wage discrimination: reduced form and structural estimates," *Journal of Human resources*, pp. 436–455.

Canning, J., M. Haque, and Y. Wang (2012): "Women at the wheel: Do female executives drive start-up success," *Dow Jones and Company.*

DiNardo, J., J. Matsudaira, J. McCrary, and L. Sanbonmatsu (2021): "A Practical Proactive Proposal for Dealing with Attrition: Alternative Approaches and an Empirical Example," *Journal of Labor Economics*, 39(S2), S507–S541.

Dutz, D., I. Huitfeldt, S. Lacouture, M. Mogstad, A. Torgovitsky, and W. Van Dijk (2021): "Selection in Surveys: Using Randomized Incentives to Detect and Account for Nonresponse Bias," Discussion paper, National Bureau of Economic Research.

Fairlie, R. W. (2005): "An extension of the Blinder-Oaxaca decomposition technique to logit and probit models," *Journal of economic and social measurement*, 30(4), 305–316.

Fisher, R. A. (1922): "On the mathematical foundations of theoretical statistics," *Philosophical transactions of the Royal Society of London. Series A, containing papers of a mathematical or physical character*, 222(594-604), 309–368.

Gneezy, U., and A. Rustichini (2000): "Pay enough or don't pay at all," *The Quarterly journal of economics*, 115(3), 791–810.

Greene, P. G., M. M. Hart, E. J. Gatewood, C. G. Brush, and N. M. Carter (2003): "Women entrepreneurs: Moving front and center: An overview of research and theory," *Coleman White Paper Series*, 3(1), 1–47.

Guzman, J., and A. O. Kacperczyk (2019): "Gender gap in entrepreneurship," *Research Policy*, 48(7), 1666–1680.

Heckman, J. J. (1979): "Sample selection bias as a specification error," *Econometrica: Journal of the econometric society*, pp. 153–161.

Lee, D. S. (2005): "Training, wages, and sample selection: Estimating sharp bounds on treatment effects," .

Molina Millán, T., and K. Macours (2017): "Attrition in randomized control trials: Using tracking information to correct bias," .

Oaxaca, R. (1973): "Male-female wage differentials in urban labor markets," *International economic review*, pp. 693–709.

Van de Ven, W. P., and B. M. Van Praag (1981): "The demand for deductibles in private health insurance: A probit model with sample selection," *Journal of econometrics*, 17(2), 229–252.