ABSTRACT

ESSAYS ON EDUCATION, HEALTH, AND MISREPORTED PROGRAM PARTICIPATION

By

ERICA CATHERINE LOUIS MTENGA

AUGUST, 2025

Committee Chair: Dr. Pierre Nguimkeu

Major Department: Economics

My dissertation examines the estimation of treatment effects in presence of plausible data quality limitations related to program participation status. In addition, my dissertation also uses a novel administrative dataset to estimate the impact of air pollution exposure, plausibly widening inequalities in testing conditions, on high-stakes exam performance in Tanzania. My dissertation aims to provide reliable estimates of the impact of the program(s) to help policymakers design cost-effective and potentially welfare-improving interventions. It also informs education and labor market policy on inequalities in high-stakes exam testing conditions due to air pollution exposure, which may add noise to this measure of student ability and lead to suboptimal education and labor market outcomes.

The first chapter proposes a method to consistently estimate the individual and joint treatment effect of overlapping (and exogenous) programs that are plausibly misreported. This chapter provides the asymptotic bias expression of the ordinary least squares (OLS) estimator and shows that it is not possible to determine the direction of the bias a priori. The joint treatment effect may also have an opposite sign to the true effect, which may have dramatic consequences if used to inform policy on whether the programs are complements or substitutes. This chapter then develops a consistent estimator of treatment effects using misclassification probabilities, which may be available through validation studies and other external sources. When misclassification probabilities are unknown, the chapter provides a two-step approach, estimating them in the first step and applying them in the proposed method in the second step. In addition, we present a way to compute the average marginal effects of participating in a single program in this framework

with measurement error in multiple binary regressors. Monte Carlo simulations show that the estimator performs well in finite samples and is superior to the naive OLS estimator. Finally, we provide an empirical example, estimating the effect of the Supplemental Nutrition Assistance Program (SNAP) and the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) on food security and healthy eating using National Household Food Acquisition and Purchase Survey (FoodAPS) data.

The second chapter proposes a method to estimate treatment effects when program participation is plausibly missing. The chapter considers endogenous participation and explores different missing data mechanisms, including missing at random (MAR) and the general case of missing not at random (MNAR). The asymptotic bias expression for complete-case analysis OLS and Instrumental Variables (IV) estimators are provided and discussed. The chapter then proposes a consistent estimator of the treatment effects, the three-step estimator. The first step reframes the missing problem to that of misreporting by assigning missing data to program non-participation status. The second step estimates the true participation status when information regarding participation and missingness is available. The last step uses the predicted participation status to obtain consistent treatment effect estimates. Next, the chapter assesses the performance of this estimator in finite samples through Monte Carlo simulations and compares it with other approaches. An empirical example, estimating the impact of maternal prenatal smoking on birth weight using U.S. Natality data, is provided to illustrate the application of the proposed method in empirical studies.

The third chapter uses novel data on students' performance on national exams administered during secondary schooling in Tanzania to study how air pollution exposure on the day of the exam affects students' performance on these exams. The chapter leverages plausibly exogenous changes in local wind direction in an IV setup to obtain causal effects. IV estimates show that an increase in $PM_{2.5}$ concentration by $10\ \mu g/m^3$ on the day a student appears for the exam worsens their performance on the exam by 0.05 standard deviations. These results are robust to a host of falsification checks. The chapter also documents that the effects are pronounced for younger students, girls, students appearing for exams in government-owned examination centers, students in poorer regions, and those at the lower end of the achievement distribution. Further, the chapter provides suggestive evidence that adverse effects of air pollution on exams that test fluid intelligence drive the main results.

ESSAYS ON EDUCATION, HEALTH, AND MISREPORTED PROGRAM PARTICIPATION

BY

ERICA CATHERINE LOUIS MTENGA

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree
of
Doctor of Philosophy
in the
Andrew Young School of Policy Studies
of
Georgia State University

GEORGIA STATE UNIVERSITY
2025

ACCEPTANCE

This dissertation was prepared under the direction of the candidate's Dissertation Committee. It has been approved and accepted by all members of that committee, and it has been accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Economics in the Andrew Young School of Policy Studies of Georgia State University.

Dissertation Chair: Dr. Pierre Nguimkeu

Committee: Dr. Alberto Chong
Dr. Augustine Denteh
Dr. Rusty Tchernis

Electronic Version Approved:

Thomas J. Vicino, Dean
Andrew Young School of Policy Studies
Georgia State University
AUGUST, 2025

Dedication

To my family, whose constant love and unwavering support made this possible.

Acknowledgments

I am incredibly grateful to many people who inspired and supported me in completing my dissertation. I could not have achieved this significant milestone without their help and encouragement.

First and foremost, I extend my deepest appreciation to my advisor, Pierre Nguimkeu, for his excellent guidance, unwavering support, and mentorship throughout the program and in the process of writing this dissertation. I cannot thank you enough for your kindness, patience, and faith in me. Your commitment to excellence, passion, and rigor has helped me become a better economist and will continue to inspire me in my personal life and professional development. I also extend my sincere gratitude to the other members of my dissertation committee, Alberto Chong, Rusty Tchernis, and Augustine Denteh, for their availability, invaluable inputs, and unparalleled support.

I am also indebted to my family for their help in achieving this milestone. I am eternally grateful to my father, Louis, my mother, Warilanga, my brother, Erick, and his wife, Joyce. Your love, sacrifices, and faith in me have been a great source of strength and motivation, and words fail me to adequately express my gratitude for all the support you have extended to me throughout this journey. I would also like to thank my partner, Jose Xilau. I am incredibly grateful for your constant love, support, and encouragement. You have made this journey more enriching, and approaching the end of our PhD studies together makes me think about and appreciate how much we have uplifted each other throughout this journey.

Finally, I would like to thank all my friends and colleagues for their valuable insights, collaboration, and friendship. I feel incredibly blessed to have you in my life.

Table of Contents

List of Tables

List of Figures

Chapter I: Estimating Treatment Effects in the Presence of Overlapping Programs and Misreporting[1]

1.1 Introduction

This paper focuses on estimating individual and joint treatment effects in parametric regressions in the presence of non-mutually exclusive interventions and misreporting. Non-mutually exclusive interventions are a common feature of programs implemented in most countries. For instance, the U.S. Department of Agriculture (USDA) administers 15 domestic food assistance programs to eligible low-income households (Oliveira, 2018). These programs vary in terms of their sizes, benefits, and population targets, and low-income households may be eligible to participate in multiple programs. We can find similar overlapping programs in developing countries, such as Conditional Cash Transfers (CCTs) and Insecticide Treated Nets (ITNs) distribution programs. The CCTs require recipients to meet established criteria, including children's school attendance, up-to-date vaccinations, and regular health care visits, which may influence household's chances of receiving free or subsidized ITNs channeled through schools and antenatal Clinics (Fiszbein, 2009; Scates et al., 2020).

Estimating individual and joint treatment effects of overlapping programs may interest researchers and policymakers. First, participating in one program may increase individuals' awareness of other available programs, affecting their probability of participating in these programs. The transactional cost of participating in a program may also change conditional on participation in other programs, which may encourage uptake. For example, automatic or categorical eligibility allows households to be automatically eligible for a program without going through eligibility determination conditional on receiving other programs. Second, understanding whether parallel treatments are complements or substitutes is critical for the optimal combination of interventions and cost-effective allocation of resources. If two programs are complements, there are additional gains from implementing both programs relative to each, incentivizing the parallel implementation of the programs. In contrast, if the two programs are substitutes, the gains

[1] This paper is a joint work with Pierre Nguimkeu (Department of Economics, Georgia State University, Email: nnguimkeu@gsu.edu).

from implementing both are less than the sum of benefits from implementing either. The optimal policy, in this case, might be to implement one of the programs but not both.

However, participation in social programs is substantially misreported in survey data. False negatives occur when program participants report not receiving treatment when they did, and false positives when program nonparticipants report receiving treatment when they didn't. In this paper, we focus on false negatives, which are more prevalent in social programs in observational studies. For instance, Meyer et al. (2022) report up to 48.98%, 33.08%, and 22.82% rates of false negatives in Supplemental Nutrition Assistance Program (SNAP) participation in the Current Population Survey (CPS), the American Community Survey (ACS), and the Survey of Income and Program Participation (SIPP), while false positive is typically low, 0.73%, 0.84%, and 1.64%. While other error sources, such as recall, salience, and design of survey instruments, may result in either false negatives or positives, stigma or social image concern is likely to drive the high prevalence of underreporting in social programs (Celhay et al., 2022). Note that underreporting in surveys is not restricted to social program participation. Stigma-related misclassifications extend to other behaviors perceived as socially undesirable. For example, in health literature, social stigma may reduce the probability that a mother reports prenatal smoking in surveys (Brachet, 2008; Fertig, 2010).

Identification and estimation with single misclassified binary regressors have been extensively studied in different settings such as exogenous misreporting and treatment (e.g., Aigner 1973, Bollinger 1996, Black et al. 2000, Lewbel 2007, Chen et al. 2008b, Chen et al. 2008a, van Hasselt and Bollinger 2012, Nguimkeu et al. 2021), exogenous misreporting and endogenous treatment selection (e.g., Kane et al. 1999, Frazis and Loewenstein 2003, Brachet 2008, DiTraglia and García-Jimeno 2017, Bollinger and van Hasselt 2017, Ura 2018) and endogenous misreporting and treatment selection (e.g., Kreider et al. 2012, Hu et al. 2015, Hu et al. 2016, Nguimkeu et al. 2019). These studies have offered various approaches to recovering consistent treatment effect point estimates and, in some cases, parameter bounds. Some solutions involve adjusting OLS using knowledge of misclassification probabilities or estimating them in the data given the distribution assumptions (e.g., Aigner 1973, Nguimkeu et al. 2021). Other studies use instrumental variables (e.g., Black et al. 2000, Frazis and Loewenstein 2003, Mahajan 2006, Ura 2018, DiTraglia and García-Jimeno 2017, Ura 2018, Nguimkeu et al. 2019), and repeated measurements (e.g., Kane et al. 1999, Black et al. 2000, Chen et al. 2008b, van Hasselt and Bollinger 2012), while others estimate parameter bounds (e.g., Bollinger 1996, Black et al. 2000).

Much less is known about estimation with overlapping and potentially misclassified binary regressors. Our study is related to Jensen et al. (2019), examining whether joint participation in Supplemental Nutrition Assistance Program (SNAP) and the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) increases food security compared to participating in only SNAP. However, the study assumes household true participation status in one program, SNAP, is observed using auxiliary administrative data to validate the self-reported status. In contrast, the second program, WIC, is potentially misreported, whereas we allow both programs to be underreported. In addition, the study identifies ATE bounds under the assumption that the better food security outcome of the household is weakly increasing with household expenditure on food at home compared to total spending. Though this assumption is plausible, it seems relatively strong. The share of food expenditure to total spending may increase as households become poorer or as household size increase for a given income which may not necessarily translate to favorable food security outcome. Moreover, this study does not produce point estimates which could be of interest in policy. Garber and Klepper (1980) attempts to show bias in multiple mismeasured continuous regressors and show that at least one coefficient estimate will be attenuated. Little is known about our setting when measurement errors, non-classical in nature, occur in multiple binary regressors.

This paper proposes a solution to consistently estimate individual and joint treatment effects in a multivariate linear regression when binary regressors representing correlated and non-mutually exclusive interventions are plausibly mismeasured. We first derive the asymptotic bias in the naive OLS estimator and show that it is not possible to determine the direction of the bias a priori. We further show that OLS estimates of the joint treatment effect of the two mismeasured binary regressors can have an opposite sign to the true effect. In addition, any valid instrument will also be correlated with the measurement errors, failing to meet relevance criteria, resulting in biased treatment effects estimates. Our approach to correct for OLS bias uses known misreporting probabilities to express the correlation between true (unobserved) binary regressors and observed covariates as a function of observed covariates and misreporting probabilities. When the misclassification probabilities are unknown, we extend the maximum likelihood estimator of Hausman et al. (1998) to bivariate models and estimate them given the data. Estimated misclassification probabilities and treatment effects are consistent, given the correct distribution assumption of the true binary regressors. We then assess the finite sample performance of the proposed estimators in Monte Carlo simulations and show that the proposed estimator is superior to naive OLS. Finally, we provide an empirical example. We examine the impact of SNAP and WIC on food security and Healthy Eating Index (HEI) using the National Household Food Acquisition and Purchase Survey (FoodAPS) data.

The rest of the paper proceeds as follows. Section 1.2 reviews the literature on measurement errors in binary regressors. Section 1.3 describes our framework, a multivariate linear regression model with two correlated misclassified binary regressors and their interaction, and assesses the bias of the OLS estimator. Section 1.4 develops our proposed estimators when misclassification probabilities are available to researchers and when they are unknown. Section 1.5 provides Monte Carlo simulations. We present an empirical example in Section 1.6 and summarize our findings in Section 1.7. Mathematical proofs are in the appendix.

1.2 Literature Review

Literature on measurement errors in binary regressors mainly focuses on identification and estimation with single misclassified binary regressors under different assumptions regarding misreporting and treatment selection. A group of studies examine when misreporting and participation happen exogenously (Aigner, 1973; Bollinger, 1996; Black et al., 2000; Lewbel, 2007; Chen et al., 2008b,a; van Hasselt and Bollinger, 2012; Nguimkeu et al., 2021). Aigner

(1973) is the first study to examine misclassification in binary regressors. The study demonstrates the attenuation bias in the OLS estimator and proposes an estimator to compute treatment effects using misclassification probabilities. Black et al. (2000) provide partial identification bounds in linear regressions such that the true value of the parameter is bounded between the OLS estimator, which can be improved further if two noisy measures exist, and the instrumental variable (IV) estimator. Nguimkeu et al. (2021), estimate linear regression models with misclassified binary regressor, potentially correlated with other regressors resulting in hidden bias. The study provides a way to consistently obtain treatment effects using misclassification probabilities in a bias-adjusted least-squares estimator (BALS). Lewbel (2007) uses an instrument for participation in nonparametric and semi-parametric regressions. Other studies take a partial identification approach and provide parameter bounds under a weak set of assumptions in linear regression models (Bollinger, 1996; van Hasselt and Bollinger, 2012), and others examine identification in nonparametric regression models (Chen et al., 2008b,a).

A related group of studies examine the case with exogenous misreporting but participation or treatment selection is endogenous (Kane et al., 1999; Frazis and Loewenstein, 2003; Brachet, 2008; DiTraglia and García-Jimeno, 2017; Bollinger and van Hasselt, 2017; Ura, 2018). Kane et al. (1999) propose a generalized method of moment (GMM) to simultaneously identify misreporting errors and parameters of interest when repeated measures are available. Brachet (2008) uses a two-step GMM procedure, estimating the probability of true status in the first stage by maximum likelihood and using the predicted probabilities to recover consistent estimates in the second stage. Mahajan (2006) provides nonparametric point estimates of homogenous average treatment effects using additional information or "instrument-like variable". Frazis and Loewenstein (2003) provide homogenous average treatment effects bounds using IV and GMM. Ura (2018) provides finite bounds of local heterogenous treatment effects in a nonparametric setting using IV. DiTraglia and García-Jimeno (2017) considers identification when a discrete-valued instrument is available. Bollinger and van Hasselt (2017) propose a Bayesian approach to identify parameter bounds.

Other studies provide solutions when a single misclassified and endogenous binary regressor is endogenously misreported. Kreider et al. (2012) shows the effect of SNAP on health outcomes by estimating average treatment effects bounds under increasingly stronger but plausible nonparametric assumptions. Hu et al. (2015) uses a local polynomial regression estimator to identify parameters in single-index models. Hu et al. (2016) examines a class of nonseparable index models with measurement errors and endogeneity. Nguimkeu et al. (2019) considers underreporting cases and shows that regardless of whether participation is endogenous, endogenous misreporting results in inconsistent OLS (may lead to sign switching) and IV estimators. The study proposes a two-step estimator that estimates the probability of true participation status in the first stage using information regarding participation and misreporting and identifies treatment effects in the second stage using the predicted participation status from the first stage.

Literature on regression frameworks with multiple misclassified binary regressors is scanty. Jensen et al. (2019) identify ATE bounds on whether joint SNAP and WIC participation reduces food insecurity compared with participating in only SNAP using a nonparametric approach. One limitation of this study is that it does not yield point estimates that could be relevant to policymakers. Additionally, it considers only one program, WIC, to be misreported, and true SNAP participation status is observed in the data. Moreover, the study imposes behavioral assumptions on the relationship between outcome, program participation, and other covariates. Other studies consider the case when continuous multiple variables are measured with errors (e.g., Garber and Klepper 1980). However, these theoretical conclusions may not necessarily extend to measurement errors in multiple binary regressors, considering the non-classical nature of misclassifications in binary regressors. Unlike Savoca (2000), we consider joint treatment effects and exploit the case when misclassification probabilities are unknown.

Our paper has three salient contributions. First, we propose a consistent estimator of treatment effects of overlapping and plausibly misreported programs when information about misclassification probabilities is available, for example, through validation studies or other relevant sources. Second, when information about misclassification probabilities is unavailable, we propose a framework to estimate them from the data. Third, we discuss the computation of average marginal effects and their corresponding asymptotic standard errors in this complicated framework, as researchers and policymakers are often interested in understanding the impact of participating in a program versus not participating in it. We also show the finite sample performance of our proposed solutions and provide an empirical example.

1.3 Framework

This section describes our regression framework and discusses the bias of the OLS estimator due to misreporting.

1.3.1 Model with Overlapping Treatments and Misreporting

Our regression framework considers a multiple linear regression model with a scalar outcome, y_i, correlated and exogenous (true) participation indicators, t_{1i}^* and t_{2i}^* such that $Pr[t_{1i}^* = 1] = P_1^*$ and $Pr[t_{2i}^* = 1] = P_2^*$ with $P_1^* \in (0,1)$ and $P_2^* \in (0,1)$, and a $k \times 1$ vector of exogenous error-free covariates, x_i for each observation i in random sample size, n. The relationship between these entities is given by

$$y_i = \alpha_1 t_{1i}^* + \alpha_2 t_{2i}^* + \alpha_3 t_{1i}^* \times t_{2i}^* + x_i'\beta + \epsilon_i \tag{1}$$

Here, α_1 is the conditional average treatment effect of t_{1i}^* for subjects that are not participating in t_{2i}^*, and α_2 is the conditional average treatment effects of t_{2i}^* for subjects that are not participating in t_{1i}^*. The parameter α_3 captures the additional effect for participating in both t_{1i}^* and t_{1i}^* compared to just participating in one of them.

We assume the two true binary regressors of interest, t_{1i}^* and t_{2i}^* are correlated with $\text{Cor}(t_{1i}^*, t_{2i}^*) = \rho$. When $\rho > 0$, individuals who participate in one program are likely to also participate in the other program, for instance, as they become categorically eligible or more aware of other programs. However, if there are additional (explicit or implicit) costs or disincentives associated with participating in the other programs given participation in the first one, we expect $\rho < 0$. When $\rho = 0$, participation in these programs are independent to one another.

For instance, we expect the correlation between SNAP and WIC, ρ, to be positive as participation of individuals or certain family members in SNAP automatically qualifies pregnant, postpartum, and breastfeeding women and families with infants and children under five for WIC. SNAP and WIC cross-program collaboration may also increase the uptake of one program conditional on participating in the other through data sharing and making referrals. In addition, in many states, SNAP (or WIC) participants are also given essential information about eligibility, the application process, and the benefits of WIC (or SNAP), which may encourage them to participate in both programs.

In the treatment effect literature, the interaction term captures the effect of one program that is influenced by participation status in the other program. In other words, it reflects cases when participation in two programs has an additional effect beyond the effects of participating in individual programs alone. When the interaction term is omitted or when $\alpha_3 = 0$, the specification imposes an additive assumption indicating that the effect of t^*_{1i} on y_i is independent of the effect of t^*_{2i}, and vice versa. Note that t^*_{1i} and t^*_{2i} could be independent to each other but still have dependent effects on the outcome y_i.

Additionally, the two programs, t^*_{1i} and t^*_{2i}, are said to be complements, $\alpha_3 > 0$, if there are additional benefits of administering both programs relative to one and substitutes, $\alpha_3 < 0$ if benefits from implementing both are less than those of implementing either of the programs. If programs are complements, then parallel implementation may be welfare-improving and cost-effective, whereas substitute programs may incentivize the implementation of one rather than both programs.

We aim to consistently estimate the model parameters, $\theta = [\alpha_1, \alpha_2, \alpha_3, \beta]'$. We are particularly interested in α_3, the key parameter capturing the additional effect of participating in both programs relative to either, which we term the 'joint effect'.

Assumption 1. $\mathbb{E}\left[\epsilon_i | t^*_{1i}, t^*_{2i}, x_i\right] = 0$

We assume that (true) program participation status, t^*_{1i} and t^*_{2i}, and other covariates, x_i, in the treatment effect model are orthogonal to the error term ϵ_i. Thus, the error term does not have any predictive power on our outcome of interest, y_i, conditional on covariates and correct model specification. This assumption is standard in linear regression models.

Assumption 2. $\text{Var}(x_i)$ *exists and is nonsingular (and finite).*

This assumption is the typical identification condition in treatment effect models. It requires a matrix $X = [x'_1, x'_2, \ldots, x'_n]'$, to have a full rank, k, ruling out perfect multicollinearity among the covariates in X. Additionally, we assume that covariates follow a well-behaved distribution with finite moments, thereby eliminating extreme cases. Under Assumption (1) and (2), OLS estimator of the model parameters, $(\alpha_1, \alpha_2, \alpha_3, \beta)$, is unbiased and consistent. In particular, the probability limit of our coefficient of interest is equal to the true value, $\text{plim}\ \hat{\alpha}_{3LS} = \alpha_3$ and we can make correct policy prescriptions based on whether the two programs are complements or substitutes.

However, the econometrician does not observe the true participation status (t^*_{1i}, t^*_{2i}), but plausibly error-driven proxies, t_{1i} and t_{2i}, such that $Pr[t_{ji} = 1] = P_j$ with $P_j \in (0, 1)$. We model misreporting in the observed participation status using two unobserved binary indicators, δ_{1i} and δ_{2i}. Each of these misreporting (or reporting) indicators is such that a respondent correctly reports their treatment status if the indicator takes the value 1, and reports not receiving treatment otherwise.

Assumption 3. *The observed (error-ridden) binary regressors, (t_{1i}, t_{2i}), are functions of the true (unobserved) binary regressors and the misreporting indicators, $(\delta_{1i}, \delta_{2i})$, such that*

$$t_{1i} = t^*_{1i}\delta_{1i} \quad \textit{and} \quad t_{2i} = t^*_{2i}\delta_{2i}$$

This specification reflects a one-sided misreporting case commonly encountered in social programs especially those associated with stigma (such as SNAP and WIC), in risky behavior (such as smoking and drunk driving), and other types of responses that are prone to social desirability bias. In this case, the observed program participants truly received the treatment, but some of the observed nonparticipants also received treatment but reported no receipt, representing the false negatives. For the SNAP and WIC programs that we use in our applications, validation studies show that the prevalence of false negatives are substantial in survey data while false positives are negligible, consistent with Assumption 3 (e.g., Meyer et al. 2022). This measurement error framework resembles the partial observability model in Poirier (1980) where each of the observed binary regressors is determined jointly by the true underlying binary regressors and measurement error indicator, and it has been previously used in the measurement error literature (e.g., see Nguimkeu et al. 2019). Note that measurement errors in binary regressors are necessarily negatively correlated with true participation status (Aigner, 1973).

We allow misreporting to depend on covariates only through the true participation status. This implies that conditional on true participation status, the probability of misreporting in both programs is exogenous, hence constant and uncorrelated with ϵ_i. Specifically, we define misreporting probabilities as

Assumption 4.

$$\Pr(t_{1i} = 0 | t^*_{1i} = 1, t_{2i}, x_i) = a_1 \quad \textit{and} \quad \Pr(t_{2i} = 0 | t^*_{2i} = 1, t_{1i}, x_i) = a_2, \quad a_1, a_2 \in [0, 1). \quad (2)$$

This assumption introduces misclassification probabilities, a_1 and a_2, the probability of false negatives in t_1 and t_2, respectively. The assumption of constant misclassification probabilities is common in the literature of mismeasured binary regressors (e.g., see Bollinger 1996, Kane et al. 1999, Hausman et al. 1998, Black et al. 2000, Frazis and Loewenstein 2003, Brachet 2008, van Hasselt and Bollinger 2012, Bollinger and van Hasselt 2017, DiTraglia and García-Jimeno 2017, Nguimkeu et al. 2021). We further assume that $a_j \in [0, 1)$, ruling out severe cases of misreporting where the observed error-driven proxies are no longer informative of the true treatment status. In other words, the misclassification errors do not overwhelm their ability to signal true status, equivalent to saying that $\mathrm{Cov}(t_{ji}, t^*_{ji}) \geq 0$, and t_{ji} is better proxy of t^*_{ji} compare to $1 - t_{1i}$ or a random guess. Hausman et al. (1998) termed this assumption the "Monotonicity condition."

The econometrician does not observe the data-generating process and estimates a model equivalent to that specified in Equation (1), referred to as the operation model, using observed treatment status contaminated by misreporting errors and other error-free covariates, given by:

$$y_i = \alpha_1 t_{1i} + \alpha_2 t_{2i} + \alpha_3 t_{1i} \times t_{2i} + x_i' \beta + \epsilon_i \tag{3}$$

Naive OLS estimator of the parameters in Equation 3 are asymptotically biased given the measurement errors in the regressors. Moreover, the misreporting errors in the two binary regressors contaminate the interaction term, which imposes additional complications in estimating the model parameters.

1.3.2 Bias of Treatment Effects due to Misreporting

We first derive the bias in a naive OLS estimator of the treatment effects. We next show the joint treatment effects estimates may attain a sign opposite of the true treatment effect. To do this, we express our covariates in vector and matrix forms and define $Z^* = [t_1^*, t_2^*, t_3^*, x]$ and $Z = [t_1, t_2, t_3, x]$ where $t_{3i}^* = t_{1i}^* \times t_{2i}^*$ and $t_{3i} = t_{3i}^* \times \delta_{3i}$, with $\delta_{3i} = \delta_{1i} \times \delta_{2i}$. The following result follows.

Lemma 1. *Under Assumptions 1-2, the probability limit of the OLS estimator of the model parameters, $\theta = [\alpha_1, \alpha_2, \alpha_3, \beta]'$, is given by*

$$plim\, \hat{\theta}_{LS} = [\text{Var(Z)}]^{-1} \text{Cov}(Z, Z^*) \theta$$

Proof. See Appendix A.1.

A naive OLS estimator is biased and inconsistent because $\text{Var}(Z)] \neq \text{Cov}(Z, Z^*)$, unless there is no misreporting. To see this, consider the components of $\text{Var}(Z)$ and $\text{Cov}(Z, Z^*)$ featured in Lemma 1. First, $\text{Cov}(t_{ji}, t_{ji}^*) \neq \text{Var}(t_{ji})$, for $j \in \{1, 2, 3\}$ since the misreporting error in binary regressors is necessarily negatively correlated with true participation status,

$\text{Cov}(t_{j1}, t^*_{j1} - t_{j1}) \neq 0$. Second, misreporting errors are necessarily correlated with other regressors in the model if the true (unobserved) binary regressor is correlated with those regressors (Nguimkeu et al., 2021). We also derive this result in our framework and show that $\text{Cov}(x_i, t^*_{ji}) \neq \text{Cov}(x_i, t_{ji})$ as long as $\text{Cov}(x_i, t^*_{ji}) \neq 0$. Other terms in Lemma 1 are driven by the correlation between the two binary regressors t_{1i} and t_{2i}, and their correlation with the interaction term, t_{3i}. Since $\rho^* \neq 0$, it follows that $\text{Cov}(t_{ki}, t^*_{ji}) \neq \text{Cov}(t_{ki}, t_{ji})$, for $k \in \{1, 2, 3\}$ and $k \neq j$. In contrast, when there is no misreporting, $t^*_{ji} = t_{ji}$ and $Z^* = Z$, and the OLS estimator is unbiased and consistent.

In particular, under Assumptions 1-2, the asymptotic bias in the OLS estimator of the joint effect, α_{3LS}, is

$$\text{plim} \quad \hat{\alpha}_{3LS} - \alpha_3 = \frac{A - B\alpha_3}{Q}$$

Where $A = \mathbb{E}\left[t_{3i}\text{z}'^*_{i,-t_{3i}}\theta_{-\alpha_3}\right] - \mathbb{E}\left[t_{3i}\text{z}'_{i,-t_{3i}}\right]\left(\mathbb{E}\left[\text{z}_{i,-t_{3i}}\text{z}'_{i,-t_{3i}}\right]\right)^{-1}\mathbb{E}\left[\text{z}_{i,-t_{3i}}\text{z}^*_{i,-t_{3i}}\theta_{-\alpha_3}\right]$,

$B = \mathbb{E}\left[t_{3i}\text{z}'_{i,-t_{3i}}\right]\left(\mathbb{E}\left[\text{z}_{i,-t_{3i}}\text{z}'_{i,-t_{3i}}\right]\right)^{-1}\mathbb{E}\left[\text{z}_{i,-t_{3i}}\left(t^*_{3i} - t_{3i}\right)\right]$,

$Q = \mathbb{E}\left[t_{3i}\right] - \mathbb{E}\left[t_{3i}\text{z}'_{i,-t_{3i}}\right]\left(\mathbb{E}\left[\text{z}_{i,-t_{3i}}\text{z}'_{-3i}\right]\right)^{-1}\mathbb{E}\left[\text{z}_{i,-t_{3i}}t_{3i}\right]$,

$\theta_{-\alpha_3} = [\alpha_1,\ \alpha_2,\ \beta]'$, and $\text{z}_{i,-t_{3i}} = [t_{1i}; t_{2i}; x_i]$

Proof. See Appendix A.1.

$A \neq 0$ and $B \neq 0$, if there is misreporting in at least one binary regressor, and Q, is always positive by Cauchy-Schwarz inequality. Therefore, the direction of the bias is driven by A and B, and we would overestimate the joint treatment effect, an expansion bias when $B > 0$ and $\alpha_3 < A/B$ or $B < 0$ and $\alpha_3 > A/B$. On the other hand, we will underestimate the joint treatment effect when $B > 0$ and $\alpha_3 > A/B$ or $B < 0$ and $\alpha_3 < A/B$. Thus, the direction of the bias cannot be determined a priori. Garber and Klepper (1980) reached a similar theoretical conclusion with multiple mismeasured continuous covariates where the study shows that their coefficients are not necessarily attenuated, but at least one of them is.

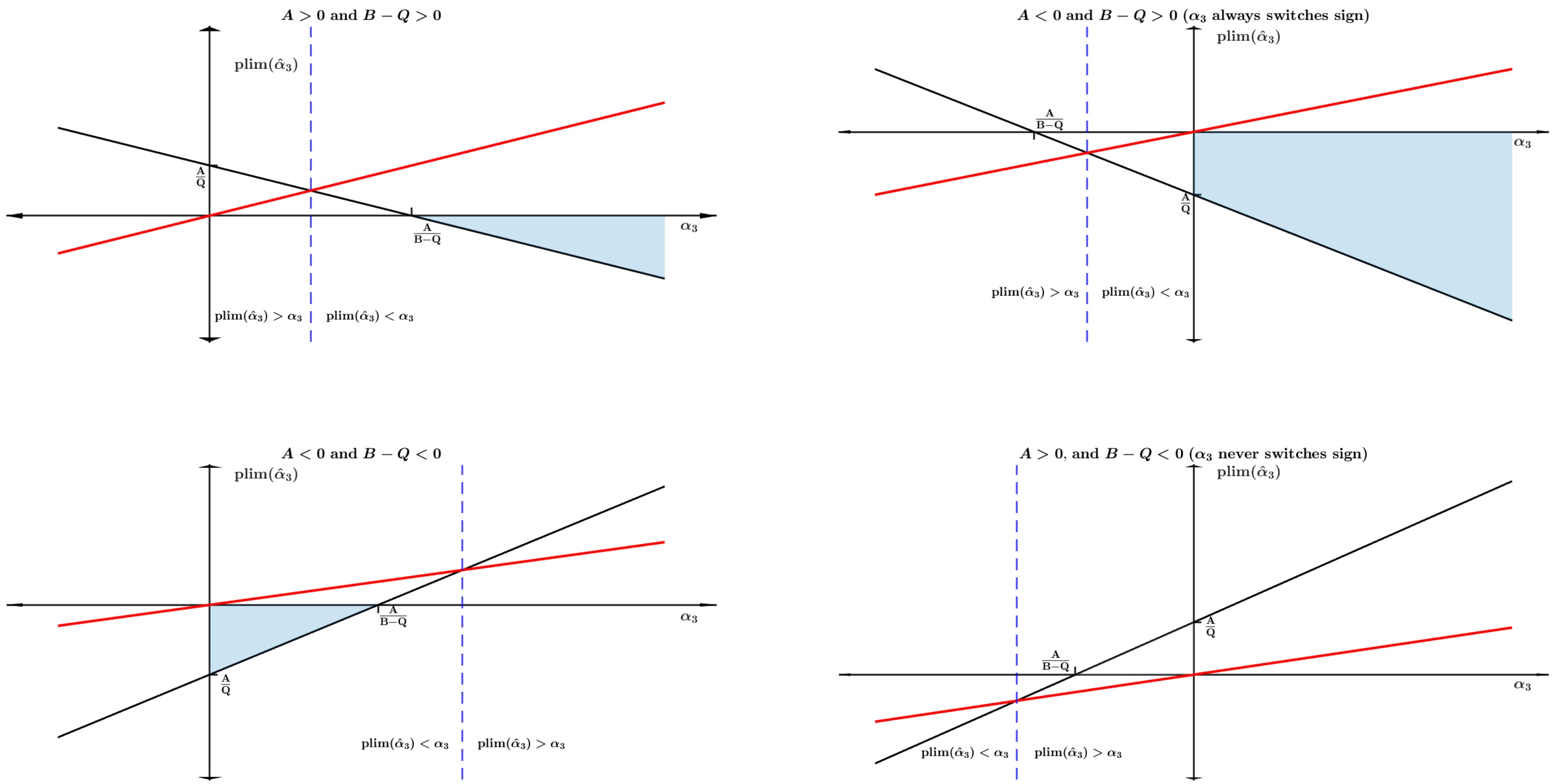
A naive OLS estimator of the joint treatment effect may also result in sign reversal. For example, with complement treatments, $\alpha_3 > 0$. However, when $A > 0$ and $B - Q > 0$, plim α_3

will always be negative. In a different setup, when $A < 0$ and $B - Q < 0$, plim α_3 will be negative when $0 < \alpha_3 < A/(B - Q)$. Without losing generality, noting that sign switching regions when $\alpha_3 < 0$ simply mirror those when $\alpha_3 > 0$, we only present the sign-switching regions when $\alpha_3 > 0$ in Figure 1.

Sign switching bias in α_3 can have dramatic welfare implications and may lead to suboptimal allocation of resources if used to inform policies regarding whether overlapping programs are complements or substitutes. Thus, for programs that are substitutes, researchers may end up falsely concluding they complement each other and are beneficial if implemented together, and vice versa. For example, in contexts where participants of both SNAP and WIC have better food security outcomes compared to participants of individual programs, sign switching could misinform policy and food assistance programs design that participating in both SNAP and WIC worsens food insecurity relative to participating in only SNAP or WIC.

When there is no misreporting, $A = 0$ and $B = 0$, naive OLS estimator is unbiased, and we have plim $\hat{\alpha}_3 = \alpha_3$. The estimated average marginal effects would also be biased when there is misreporting, especially for those who participated in both programs. On average, the estimated average marginal effects of t_1 and t_2 would be less or greater than $\alpha 1$ and α_2 depending on the magnitude and direction of the parameters, $\hat{\alpha}_1$, $\hat{\alpha}_2$, and $\hat{\alpha}_3$, and the resulting bias in these coefficients. For example, if participating in t_1 and both programs has a positive impact on the outcome, the estimated average marginal effect of t_1 would be greater than α_1 if there is no sign switching in α_3, and less than α_1 if sign switching occurs.

Figure 1. Illustration of the Sign-switching Regions in OLS Estimation of the Joint Treatment Effect, $\alpha_3 > 0$



1.4 The Proposed Estimator

We propose an estimator that uses misclassification probabilities to correct for the OLS bias and attain consistent treatment effects. We first assume that econometricians or researchers have access to information regarding the misclassification probabilities. We also develop a procedure to estimate misclassification probabilities given the data and distributional assumptions about the true binary regressors.

1.4.1 Known Misclassification Probabilities

By inverting the relationship in Lemma 1, a consistent estimator of the model parameters can be obtained by defining an adjusted least squares estimator given by

$$\widehat{\theta}_{Adj} = \widehat{\text{Cov}}(Z, Z^{*})^{-1}\widehat{\text{Var}}(Z)\widehat{\theta}_{LS}. \tag{4}$$

While $\widehat{\text{Var}}(Z)$ can be obtained from the sample, $\widehat{\text{Cov}}(Z, Z^{*})$ on the other hand, the covariance between the observed regressors and their true (unobserved) counterparts, contains components that are not directly observed in the data. Our proposed estimator uses the assumptions made above for this framework to estimate $\widehat{\text{Cov}}(Z, Z^{*})$ from the data by expressing this quantity as a function of misclassification probabilities and sample statistics that can be computed from the data. This is provided by the following lemma.

Lemma 2. *Under Assumptions 1-2,*

$$\begin{aligned}
&\text{Cov}(t_{2i}, t_{1i}^{*}) = \zeta_1 \text{Cov}(t_{2i}, t_{1i}), && \text{Cov}(t_{3i}, t_{1i}^{*}) = \zeta_1 \text{Cov}(t_{1i}, t_{2i}),\\
&\text{Cov}(x_i, t_{1i}^{*}) = \zeta_1 \text{Cov}(x_i, t_{1i}), && \text{Cov}(t_{3i}, t_{2i}^{*}) = \zeta_2 \text{Cov}(t_{2i}, t_{1i}),\\
&\text{Cov}(t_{1i}, t_{2i}^{*}) = \zeta_2 \text{Cov}(t_{1i}, t_{2i}), && \text{Cov}(x_i, t_{2i}^{*}) = \zeta_2 \text{Cov}(x_i, t_{2i})\\
&\text{Cov}(t_{1i}, t_{1i}^{*}) = \eta_1 \text{Var}(t_{1i}) && \text{Cov}(t_{2i}, t_{2i}^{*}) = \eta_2 \text{Var}(t_{2i})\\
&\text{Cov}(t_{1i}, t_{3i}^{*}) = \zeta_2 \eta_{1i} \text{Cov}(t_{1i}, t_{3i}), && \text{Cov}(t_{2i}, t_{3i}^{*}) = \zeta_1 \eta_2 \text{Cov}(t_{2i}, t_{3i}),\\
&\text{Cov}(x_i, t_{3i}^{*}) = \zeta_3 \text{Cov}(x_i, t_{3i}), && \text{Cov}(t_{3i}, t_{3i}^{*}) = \eta_3 \text{Var}(t_{3i})
\end{aligned}$$

$Where$

$$\zeta_1 = \frac{1}{1-a_1}, \qquad \zeta_2 = \frac{1}{1-a_2}, \qquad \zeta_3 = \frac{1}{1-a_3},$$

$$\eta_1 = \frac{1-a_1-P_1}{(1-a_1)(1-P_1)} \qquad \eta_2 = \frac{1-a_2-P_2}{(1-a_2)(1-P_2)} \qquad \eta_3 = \frac{1-a_3-P_3}{(1-a_3)(1-P_3)},$$

$$a_3 = \Pr(t_{3i} = 0 | x_i, t_{3i}^* = 1) \quad P_3 = \Pr(t_{3i} = 1)$$

Using Lemma 2, we now have an estimator of $\widehat{\text{Cov}}(Z, Z^*)$ from the data, which we denote $\widehat{W}(a_1, a_2)$ to express its dependency to misclassification probabilities only, and we can formally present our proposed estimator of the parameters in the model given by Equation 3. The following result follows if we denote the sample covariance between r_i and s_i by σ_{rs} and variance of r_i by σ_r^2. Specifically, $\sigma_{rs} = \frac{1}{n}\sum_{i=1}^{n}(r_i - \bar{r})(s_i - \bar{s})'$ and $\sigma_r^2 = \frac{1}{n}\sum_{i=1}^{n}(r_i - \bar{r})(r_i - \bar{r})'$where $\bar{r}$ and $\bar{s}$ are the sample means of r_i and s_i.

Theorem 1. *Under Assumptions 1-2, for given misclassification probabilities, a_1 and a_2, the adjusted least squares estimator is given by*

$$\begin{bmatrix} \hat{\alpha}_{1Adj} \\ \hat{\alpha}_{2Adj} \\ \hat{\alpha}_{3Adj} \\ \hat{\beta}_{Adj} \end{bmatrix} = \begin{bmatrix} \eta_1\sigma_{t_1}^2 & \zeta_2\sigma_{t_1t_2} & \eta_1\zeta_2\sigma_{t_1t_3} & \sigma_{t_1x} \\ \zeta_1\sigma_{t_2t_1} & \eta_2\sigma_{t_2}^2 & \eta_2\zeta_1\sigma_{t_2t_3} & \sigma_{t_2x} \\ \eta_1\sigma_{t_3t_1} & \eta_2\sigma_{t_3t_2} & \eta_3\sigma_{t_3}^2 & \sigma_{t_3x} \\ \zeta_1\sigma_{xt_1} & \zeta_2\sigma_{xt_2} & \zeta_3\sigma_{xt_3} & \sigma_x^2 \end{bmatrix}^{-1} \begin{bmatrix} \sigma_{yt_1} \\ \sigma_{yt_2} \\ \sigma_{yt_3} \\ \sigma_{yx} \end{bmatrix}$$

Where ζ_j and η_j for $j \in \{1, 2, 3\}$ are as defined in Lemma 2. It follows that:

(i) These estimators are consistent, that is, $\hat{\theta}_{Adj} \xrightarrow{p} \theta$, with $\theta = [\alpha_1, \alpha_2, \alpha_3, \beta]'$

When there is no misreporting, $\eta_j = 1$ and $\zeta_j = 1$, our estimator in Theorem 1 is quantitively similar to the OLS, and both are consistent. We can also see this by expressing the adjusted least squares estimator as $\hat{\theta}_{Adj} = \widehat{W}(a_1, a_2)^{-1}\widehat{\text{Var}}(Z)\hat{\theta}_{LS}$. When there is no misreporting, i.e. $a_1 = a_2 = 0$, then $\widehat{W}(0,0) = \widehat{\text{Var}}(Z)$ and $\hat{\theta}_{Adj} = \hat{\theta}_{LS}$. When there is misclassification in at least one treatment status, solutions proposed for single misclassified regressors (e.g., modified least square estimator in Aigner (1973) and bias-adjusted least square in Nguimkeu et al. (2021)) do not account for the bias driven by misreporting in the second binary regressor. In contrast, the proposed estimator, the adjusted least squares, is consistent when misreporting exists in any or all binary regressors.

The variance of the proposed estimator can be estimated through bootstrapping techniques (Efron, 1979; Efron and Tibshirani, 1994; Brownstone and Kazimi, 1998; Horowitz, 2001). Unless the researcher knows or makes assumptions on the distribution of ϵ_i, we suggest to use a non-parametric bootstrap procedure that jointly samples the outcome, y_i, program participation status (t_{1i}, t_{2i}), and a vector of other covariates, x_i.

In our setting, the bootstrap procedure to estimate the covariance of our consistent estimator would involve M replications, samples n_m observations with replacement in each replication, m, and compute coefficient estimates using the adjusted least squares, $\hat{\theta}_{m,Adj}$ with each sample. Let $\bar{\hat{\theta}}_{M,Adj} = \frac{1}{M}\sum_{m=1}^{M}\left[\hat{\theta}_{m,Adj}\right]$, the average of the bootstrapped estimates of θ, the estimated asymptotic covariance matrix of the adjusted least squares estimator, $\hat{\theta}_{Adj}$, follows as

$$\hat{\text{Var}}\left[\hat{\theta}_{Adj}\right] = \frac{1}{M-1}\sum_{m=1}^{M}\left[\hat{\theta}_{m,Adj} - \bar{\hat{\theta}}_{M,Adj}\right]\left[\hat{\theta}_{m,Adj} - \bar{\hat{\theta}}_{M,Adj}\right]'$$

1.4.2 Estimation of Misclassification Probabilities

Our estimator so far uses known misreporting probabilities, a_1 and a_2, to estimate the treatment effects of interest. Hence, applying the proposed estimator in correcting the bias due to measurement error may be limited in the context where misclassification probabilities are unknown. One way of addressing this limitation is by estimating the misclassification probabilities in the data in the first step, $\hat{a}_1$ and $\hat{a}_2$, and using them in the proposed adjusted least squares estimator in the second step, $\hat{\theta}_{Adj} = \hat{W}(\hat{a}_1, \hat{a}_2)^{-1}\widehat{\text{Var}}(Z)\hat{\theta}_{LS}$.

The existing literature has taken this approach mainly in the context of single misreported binary regressors (e.g., Brachet 2008; Nguimkeu et al. 2021) and use the framework of estimating misclassification probabilities proposed by Hausman et al. (1998).

We can also extend the approach in Hausman et al. (1998) to a bivariate or multivariate framework to fit the context of multiple misclassified binary regressors. We model (true) binary regressors as

$$\begin{aligned} t_{1i}^* &= 1(x_i'\gamma_1 + u_{1i} > 0) \\ t_{2i}^* &= 1(x_i'\gamma_2 + u_{2i} > 0) \end{aligned} \tag{5}$$

and their interaction as

$$t_{3i}^* = 1(x_i'\gamma_1 + u_{1i} > 0, x_i'\gamma_2 + u_{2i} > 0) \tag{6}$$

We assume that the joint Cumulative Distribution Function (CDF) of $(-u_{1i}, -u_{2i})$ is known and defined by $F\left(x_i'\gamma_1, x_i'\gamma_2, \rho^*\right) = \Pr\left[-u_{1i} \le x_i'\gamma_1, -u_{2i} \le x_i'\gamma_2\right]$. In particular, if we assume that conditional on x_i, the disturbance terms (u_1, u_2) are drawn from bivariate normal distribution given by

$$\begin{pmatrix} u_{1i} \\ u_{2i} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho^* \\ \rho^* & 1 \end{pmatrix}\right), \tag{7}$$

their joint CDF would then be

$$\Pr\left[-u_{1i} \le x_i'\gamma_1, -u_{2i} \le x_i'\gamma_2\right] = \Phi_2(x_i'\gamma_1, x_i'\gamma_2; \rho^*),$$

where $\Phi_2(\cdot, \cdot, \rho^*)$ is the bivariate standard normal CDF associated with correlation coefficient ρ^*.

For given misclassification probabilities (a_1, a_2), the marginal likelihood of observed binary regressors, t_{1i} and t_{2i}, are given by

$$\begin{aligned} \Pr\left[t_{1i} = 1 | x_i\right] &= (1 - a_1)\Phi(x_i'\gamma_1) \\ \Pr\left[t_{2i} = 1 | x_i\right] &= (1 - a_2)\Phi(x_i'\gamma_2) \end{aligned} \tag{8}$$

and their joint probability can be obtained by

$$\Pr\left[t_{1i} = 1, t_{2i} = 1 | x_i\right] = (1 - a_1 - a_2 + a_1 \times a_2)\Phi_2(x_i'\gamma_1, x_i'\gamma_2, \rho^*), \tag{9}$$

where $\Phi(\cdot)$ is the univariate standard normal CDF and $\Theta = (a_1, a_2, \gamma_1, \gamma_2, \rho)$ are vectors of unknown parameters. These unknown parameters can be estimated jointly through maximum likelihood method. The maximum likelihood estimators of Θ, denoted by $\hat{\Theta}$, can be obtained by maximizing the log-likelihood function given by

$$\mathcal{L}(\Theta) = \frac{1}{n}\sum_{i=1}^{n}\{t_{1i}t_{2i}\ln\omega_{i11}(\Theta) + t_{1i}\,(1-t_{2i})\ln\omega_{i10}(\Theta) \\ + (1-t_{1i})t_{2i}\ln\omega_{i01}(\Theta) + (1-t_{1i})(1-t_{2i})\ln\omega_{i00}(\Theta)\} \quad (10)$$

Where $\omega_{i11} = \Pr(t_{1i} = 1, t_{2i} = 1)$, $\omega_{i10} = \Pr(t_{1i} = 1, t_{2i} = 0)$, $\omega_{i01} = \Pr(t_{1i} = 0, t_{2i} = 1)$, and $\omega_{i00} = \Pr(t_{1i} = 0, t_{2i} = 0)$ are probabilities responding to the four possible realizations of t_1 and t_2, defined as

$$\omega_{i11}(\Theta) = (1 - a_1 - a_2 + a_1 \times a_2)\,\Phi_2\left(x_i'\gamma_1, x_i'\gamma_2; \rho^*\right)$$

$$\omega_{i10}(\Theta) = (1 - a_1)\,\Phi\left(x_i'\gamma_1\right) - (1 - a_1 - a_2 + a_1 \times a_2)\,\Phi_2\left(x_i'\gamma_1, x_i'\gamma_2; \rho^*\right)$$

$$\omega_{i01}(\Theta) = (1 - a_2)\,\Phi\left(x_i'\gamma_2\right) - (1 - a_1 - a_2 + a_1 \times a_2)\,\Phi_2\left(x_i'\gamma_1, x_i'\gamma_2; \rho^*\right)$$

$$\omega_{i00}(\Theta) = 1 - (1 - a_1)\,\Phi\left(x_i'\gamma_1\right) - (1 - a_2)\,\Phi\left(x_i'\gamma_2\right) + \\ (1 - a_1 - a_2 + a_1 \times a_2)\,\Phi_2\left(x_i'\gamma_1, x_i'\gamma_2; \rho^*\right)$$

The maximum likelihood estimators, $\hat{\Theta}$, which includes estimators of misclassification probabilities, $\hat{a}_1$ and $\hat{a}_2$ can be obtained by maximizing likelihood function given by Equation 10 with respect Θ. Standard errors can be obtained, as usual, by computing the inverse of the observed information matrix. Given the model assumptions and correct specification of the cumulative distribution function, these maximum likelihood estimators are consistent.

1.4.3 Average Marginal Effects

Researchers and policymakers are often interested in the effect of participating in one program relative to not participating, that is, the average marginal effect of treatment 1 or 2. In our setting, as specified in Equation 1, the average marginal effects are given by

$$AME_j = \mathbb{E}_{t^*_{3-j}}\left[\frac{\partial \mathbb{E}[y_i|t^*_{1i}, t^*_{2i}, x_i]}{\partial t^*_j}\right] = \mathbb{E}_{t^*_{3-j}}\left[\alpha_j + \alpha_3 t^*_{(3-j)i}\right] = \alpha_j + \alpha_3 \frac{\mathbb{E}\left[t_{(3-j)i}\right]}{1 - a_{3-j}} \quad (11)$$

$$for \quad j \in \{1, 2\}$$

We can recover consistent estimates of the average marginal effects by using the adjusted least squares estimator to obtain coefficient estimates in the treatment effect model, $(\hat{\alpha}_{1Adj}, \hat{\alpha}_{2Adj}, \hat{\alpha}_{3Adj})$ and the misclassification probabilities which may be available through validation datasets or estimated using the framework described in Section 1.4.2. Specifically, the estimation of average marginal effects, $\widehat{AME}$, follows as

$$\widehat{AME}_j = \mathbb{E}_{\hat{t}^*_{3-j}}\left[\frac{\partial \mathbb{E}[y_i|\hat{t}^*_{1i}, \hat{t}^*_{2i}, x_i]}{\partial \hat{t}^*_j}\right] = \hat{\alpha}_{j,Adj} + \frac{\hat{\alpha}_{3,Adj}}{1 - \tilde{a}_{3-j}} \bar{t}_{3-j} \qquad for \quad j \in \{1, 2\}$$

where $\bar{t}_{3-j} = \frac{1}{n}\sum_{i=1}^{n} \hat{t}_{(3-j)i,Adj}$ and $\tilde{a}_{3-j} = a_{3-j}$ or $\tilde{a}_{3-j} = \hat{a}_{3-j}$ depending on whether the misreporting rates are known or not. By applying the delta method, we can in turn estimate the asymptotic variance of the average marginal effects as

$$\widehat{\text{Var}}(\widehat{AME}_j) = \Pi_j \widehat{\text{Var}}[\hat{\theta}_{Adj}]\Pi_j' \qquad for \quad j \in \{1, 2\}$$

where $\Pi_j = [1 \;\; 0 \;\; \frac{\bar{t}_{3-j}}{1-a_{3-j}} \;\; 0_{k\times 1}]$, a $(k+3) \times 1$ vector obtained from $\frac{\partial \widehat{AME}_j}{\partial \hat{\theta}_{Adj}}$. This approach provides us with standard errors and confidence interval required in making inference on the average marginal effects.

1.4.4 Generalization to Multivariate Binary Variables

Our proposed estimator can be extended to estimate individual and joint treatment effects of multiple and plausibly misreported binary variables. This framework is useful in the context of multiple programs where policymakers may be interested in understanding the joint effect of a pair of programs on outcomes of interest. Suppose the researcher is interested in evaluating the impact of p programs and considers potential pairwise interactions. There would be $\frac{p(p-1)}{2}$ such

interactions terms which can be denoted $t_{I1}, \ldots, t_{I\frac{p(p-1)}{2}}$. The adjusted least square estimator of the $\frac{p(p+1)}{2} + k$ model parameters in this case,
$\hat{\theta}_{Adj} = (\hat{\alpha}_{1Adj}, \ldots, \hat{\alpha}_{pAdj}, \hat{\alpha}_{I1Adj}, \ldots, \hat{\alpha}_{I\frac{p(p-1)}{2}Adj}, \hat{\beta}_{Adj})$ is given by
$\hat{\hat{\theta}}_{Adj} = \widehat{W}(a_1, \ldots, a_p)^{-1}\widehat{\mathrm{Var}}(Z)\hat{\theta}_{LS}$, such that $\widehat{W}(a_1, \ldots, a_p) = \widehat{\mathrm{Cov}}(Z, Z^*)$, and
$Z = (t_1, \ldots, t_p, t_{I1}, \ldots, t_{I\frac{p(p-1)}{2}}, x)$. Here, the terms in $\widehat{W}(a_1, \ldots, a_p)$ are obtained as for the binary case by applying the relationships of the types given in Lemma 2.

1.5 Monte Carlo Simulations

In this section, we examine the finite sample performance of the proposed estimator and compare the proposed methods with OLS through Monte Carlo simulations. Our goal is to consistently estimate the parameters (α_1, α_2, α_3, β) of the model presented in Equation 1. Specifically, we aim to consistently estimate the individual treatment effect and the joint effect of participating in both programs relative to participating in only one program, under the assumption that the true participation status of both programs, t_1^* and t_2^*, are unobserved, but only, t_1 and t_2, their error-ridden surrogates driven, are observed. In addition, we assume that other observed covariates, x_i, are correctly measured.

1.5.1 Simulation Setup

The data-generating process is simulated as follows. The true treatment indicators, t_1^* and t_2^*, are given by

$$t_{1i}^* = 1(\gamma_1 x_i + u_{1i} > 0) \text{ and } t_{2i}^* = 1(\gamma_1 x_i + u_{1i} > 0),$$

where $\gamma_1 = \gamma_2 = 1$ and $\begin{pmatrix} u_{1i} \\ u_{2i} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho^* \\ \rho^* & 1 \end{pmatrix}\right)$, $\rho^* = 0.2$. This setting can be extended to explore different degrees of correlation between the two treatment regressors by varying the values of ρ^*. The outcome equation y_i is given by

$$y_i = c + \alpha_1 t_{1i}^* + \alpha_2 t_{2i}^* + \alpha_3 t_{1i}^* \times t_{2i}^* + x_i\beta + \epsilon_i, \quad \text{where } \epsilon_i \sim N(0,1)$$

and $x_i \sim N(0, 1)$. The true population regression parameters are $c = 1$, $\alpha_1 = 5.0$, $\alpha_2 = 2.8$, $\alpha_3 = 0.3$, and $\beta = 0.5$. We aim to estimate the model parameters consistently and of most interest, the individual (conditional) average treatment effects, α_1 and α_2, and the joint average treatment effect, α_3.

The econometrician does not observe the data-generating process and (true) treatment regressors defined above, but only their erroneous surrogates. The econometrician then estimates the following operational model:

$$y_i = c + \alpha_1 t_{1i} + \alpha_2 t_{2i} + \alpha_3 t_{1i} \times t_{2i} + x_i \beta + \varepsilon_i$$

Where t_{1i} and t_2 are error-driven proxies of true treatment status defined by:

$$t_{1i} = t_{1i}^* 1(\nu_{1i} > a_1)$$

$$t_{2i} = t_{2i}^* 1(\nu_{2i} > a_2)$$

The disturbance terms, $\nu_{1i} \sim \mathcal{U}(0, 1)$ and $\nu_{2i} \sim \mathcal{U}(0, 1)$, are drawn from uniform distribution and the parameters, a_1, and a_2, are the misreporting probabilities in t_{1i} and t_{2i}, such that $a_j \in [0, 1)$ for $j \in \{1, 2\}$. The misreporting probabilities determine the proportion of false negatives in observed binary regressors. For example, if there are no false negatives, $a_j = 0$, then $1(\nu_{ji} > 0) = 1$ and $t_{ji} = t_{ji}^*$. When $a_j \neq 0$, we allow for $a_j \times 100\%$, for $j \in \{1, 2\}$ rate of false negatives. Specifically, we consider the following set of misclassification probabilities, $(a_1, a_2) \in \{(0, 0.1); (0, 0.2); (0, 0.4); (0.1, 0.1); (0.1, 0.2); (0.1, 0.4); (0.2, 0.1); (0.2, 0.2); (0.2, 0.4); (0.4, 0.1); (0.4, 0.2); (0.4, 0.4)\}$. In this way, we can examine OLS bias and the performance of the proposed estimator at different degrees of misreporting within and across programs.

We first estimate the model parameters, c, α_1, α_2, α_3, and β, using the OLS estimator and true treatment status, unobserved to the econometrician. We next report naive OLS estimates based on the observed misclassified binary regressors. We extend our OLS analysis to the problem of misspecification in the operational model, where we subsequently omit the interaction term, $t_{1i} \times t_{2i}$, and the second treatment, t_{2i} to mimic common treatment effect model specifications in empirical studies and highlight the resulting bias in naive OLS estimation. We then use our proposed estimator to estimate the model parameters. Note that our estimator uses known misclassification probabilities, a_1 and a_2. Finally, we present estimates when misclassification probabilities are unknown. We estimate them from the simulated data using the framework proposed in Section 1.4.2.

1.5.2 Simulation Results

In our Monte Carlo simulations, we execute 1000 replications using the sample size of $n = 5000$ observations. We report the averaged simulation results in Table 1 for the OLS and adjusted least squares estimators. "OLS True" columns show OLS estimates using true (unobserved) binary regressors. Our benchmark is the first column of "OLS True", which uses true binary regressors and has the correct model specification. Columns with "OLS Observed" present naive OLS estimates obtained using the observed data. Our simulation results are consistent with the theoretical discussion that misreporting errors may reverse the sign of the joint treatment effect. This sign switching can occur even at low misreporting rates, such as in cases where there is no misreporting in the first treatment and only a 10% false negative rate in the second treatment. The bias in the joint treatment effect worsens with increasing misreporting rates within and across programs. Misspecification introduces omitted variable bias, exacerbating the OLS bias in individual treatment effect estimates in some settings. Empirical studies may be affected by these inconsistencies, leading to policy prescriptions that are far from optimal.

We present the results of our proposed estimator, adjusted least squares, in the last three columns of Table 1. We first show the results when the interaction term is excluded from the model. As expected, these estimates are inconsistent due to omitted variable bias. However, these estimates are less biased than those from naive OLS, especially when misreporting in one or both programs is high. In contrast, when the model is correctly specified, the adjusted least squares estimator consistently estimates the treatment effects and other model parameters using known misreporting rates and when misreporting rates are unknown (and estimated from the data). Our estimates of the misreporting rates are also very close to the true misreporting rates in the

Table 1. Monte Carlo Simulation Results

a_1	a_2	Para.	True Values	OLS True	OLS Observed	OLS True	OLS Observed	OLS True	OLS Observed	Adj.LS - Known Misreporting Rates		Adj.LS - Unknown Misreporting Rates
		α_1	5.0	5.000	5.323	5.151	5.251	5.548	5.548	5.152	5.000	5.013
		α_2	2.8	2.800	2.550	2.951	2.461			2.951	2.800	2.781
	0.1	α_3	0.3	0.302	-0.163						0.303	0.330
		β	0.5	0.499	0.677	0.499	0.679	1.220	1.220	0.499	0.499	0.489
		c	1.0	0.998	1.247	0.954	1.272	2.231	2.231	0.953	0.998	0.992
		α_1	5.0	5.000	5.484	5.150	5.318	5.543	5.543	5.149	4.998	5.012
		α_2	2.8	2.801	2.361	2.950	2.111			2.951	2.801	2.781
0.0	0.2	α_3	0.3	0.299	-0.425						0.302	0.331
		β	0.5	0.499	0.800	0.499	0.808	1.220	1.220	0.499	0.499	0.488
		c	1.0	1.000	1.440	0.956	1.503	2.235	2.235	0.956	1.000	0.994
		α_1	5.0	5.001	5.611	5.151	5.414	5.547	5.547	5.150	4.996	5.008
		α_2	2.8	2.799	2.075	2.949	1.641			2.948	2.796	2.779
	0.4	α_3	0.3	0.301	-0.680						0.307	0.337
		β	0.5	0.500	0.967	0.500	0.981	1.221	1.221	0.501	0.501	0.489
		c	1.0	1.000	1.722	0.955	1.806	2.232	2.232	0.957	1.001	0.995
		α_1	5.0	5.000	4.790	5.149	4.383	5.542	4.644	5.150	4.999	5.000
		α_2	2.8	2.802	3.015	2.951	2.607			2.952	2.801	2.799
	0.1	α_3	0.3	0.299	-0.830						0.302	0.308
		β	0.5	0.499	0.992	0.499	1.010	1.221	1.605	0.499	0.498	0.496
		c	1.0	1.000	1.729	0.955	1.859	2.234	2.914	0.954	0.998	0.998
		α_1	5.0	5.000	4.882	5.150	4.447	5.544	4.648	5.153	4.998	5.000
		α_2	2.8	2.801	2.764	2.951	2.237			2.949	2.795	2.792
0.1	0.2	α_3	0.3	0.300	-1.004						0.310	0.314
		β	0.5	0.499	1.126	0.499	1.151	1.221	1.605	0.499	0.498	0.497
		c	1.0	1.000	1.958	0.955	2.110	2.234	2.914	0.954	0.999	1.000
		α_1	5.0	5.001	4.898	5.151	4.530	5.545	4.647	5.154	5.000	5.007
		α_2	2.8	2.798	2.393	2.948	1.744			2.957	2.804	2.798
	0.4	α_3	0.3	0.301	-1.140						0.310	0.304
		β	0.5	0.501	1.310	0.501	1.339	1.220	1.605	0.496	0.495	0.496
		c	1.0	1.000	2.298	0.955	2.442	2.232	2.913	0.947	0.991	0.993
		α_1	5.0	4.999	4.383	5.149	3.763	5.542	3.997	5.157	5.001	4.996
		α_2	2.8	2.799	3.220	2.950	2.708			2.946	2.789	2.797
	0.1	α_3	0.3	0.301	-1.182						0.313	0.307
		β	0.5	0.500	1.218	0.500	1.248	1.221	1.882	0.498	0.498	0.499
		c	1.0	1.000	2.103	0.956	2.282	2.235	3.407	0.954	0.999	1.001
		α_1	5.0	5.000	4.423	5.150	3.817	5.545	3.997	5.149	4.998	5.003
		α_2	2.8	2.800	2.934	2.949	2.329			2.951	2.801	2.798
0.2	0.2	α_3	0.3	0.299	-1.305						0.302	0.304
		β	0.5	0.500	1.361	0.500	1.398	1.220	1.883	0.501	0.500	0.498
		c	1.0	1.001	2.358	0.957	2.549	2.234	3.408	0.957	1.000	1.000
		α_1	5.0	4.999	4.359	5.151	3.894	5.547	3.999	5.155	5.000	5.001
		α_2	2.8	2.799	2.498	2.950	1.818			2.953	2.799	2.800
	0.4	α_3	0.3	0.303	-1.346						0.312	0.316
		β	0.5	0.500	1.561	0.500	1.599	1.220	1.882	0.497	0.495	0.492
		c	1.0	1.000	2.740	0.955	2.903	2.232	3.406	0.952	0.995	0.995
		α_1	5.0	5.001	3.788	5.151	2.934	5.545	3.128	5.163	5.011	5.006
		α_2	2.8	2.798	3.330	2.948	2.848			2.944	2.788	2.797
	0.1	α_3	0.3	0.300	-1.496						0.310	0.299
		β	0.5	0.501	1.529	0.501	1.567	1.222	2.258	0.500	0.498	0.501
		c	1.0	1.000	2.657	0.956	2.845	2.233	4.069	0.951	0.995	0.999
		α_1	5.0	5.002	3.752	5.151	2.984	5.544	3.132	5.176	5.011	5.005
		α_2	2.8	2.800	2.972	2.950	2.446			2.940	2.772	2.785
0.4	0.2	α_3	0.3	0.299	-1.519						0.336	0.315
		β	0.5	0.499	1.684	0.499	1.727	1.220	2.254	0.492	0.490	0.497
		c	1.0	0.999	2.947	0.955	3.131	2.234	4.065	0.943	0.989	0.996
		α_1	5.0	4.998	3.588	5.148	3.038	5.545	3.126	5.162	4.992	4.998
		α_2	2.8	2.800	2.461	2.950	1.912			2.950	2.780	2.784
	0.4	α_3	0.3	0.300	-1.453						0.347	0.327
		β	0.5	0.501	1.910	0.501	1.948	1.221	2.257	0.494	0.488	0.494
		c	1.0	1.001	3.376	0.957	3.521	2.234	4.068	0.947	0.989	0.996

data-generating process, and we report the corresponding average marginal effect in Table 2. Overall, our proposed estimator performs similarly to our benchmark, the OLS estimator, when true (unobserved and error-free) binary regressors are used, and it is superior to naive OLS.

Table 2. Monte Carlo Simulation Results: Average Marginal Effects

a_1	a_2	OLS True	OLS Observed	Adj.LS known Misreporting Rates	Adj.LS - Unknown Misreporting Rates
0.0	0.1	5.150 2.950	5.383 2.263	5.147 2.953	5.154 2.950
	0.2	5.150 2.950	5.514 1.914	5.146 2.957	5.153 2.952
	0.4	5.150 2.950	5.667 1.543	5.142 2.971	5.151 2.963
0.1	0.1	5.150 2.950	4.195 2.680	5.150 2.952	5.145 2.953
	0.2	5.150 2.950	4.311 2.306	5.159 2.951	5.156 2.950
	0.4	5.150 2.950	4.409 1.874	5.144 2.981	5.151 2.960
0.2	0.1	5.150 2.950	3.565 2.882	5.175 2.950	5.157 2.954
	0.2	5.150 2.950	3.636 2.479	5.160 2.958	5.161 2.960
	0.4	5.150 2.950	3.677 1.982	5.157 2.972	5.166 2.955
0.4	0.1	5.150 2.950	2.842 3.080	5.174 2.952	5.164 2.955
	0.2	5.150 2.950	2.851 2.599	5.196 2.943	5.163 2.953
	0.4	5.150 2.950	2.806 2.014	5.170 2.961	5.174 2.949

We next assess the sensitivity of our proposed solution to various forms of misspecification. First, we include false positives in our data-generating process by redefining the observed treatment status as $t_{ji} = t_{ji}^{*}1(\nu_{ji} > a_j) + (1 - t_{ji}^{*})1(\nu_{ji} < b_j)$ for $j \in (1,2)$. Thus, the observed treatment indicator, t_j, has ($a_j \times 100$) rate of false negatives and ($b_j \times 100$) rate of false positive when $a_j > 0$ and $b_j > 0$. We consider 1%, 5%, and 10% rates of false positives in treatment one indicator and 1% and 5% in treatment one and two indicators. The results show that our proposed estimator performs relatively well when the false positive rate is low. The estimation bias worsens with higher rates of false positives and an increase in false negatives for a given rate of false positives. Further, the bias of the interaction term is relatively lower when misclassification probabilities are estimated rather than taken as given. Second, we examine the robustness of the proposed solution to endogeneity in treatment selection or program participation by allowing correlation between u_{ji} and ϵ_i for $j \in (1, 2)$. Specifically, we consider (0.2, 0.2), (-0.2,-0.2) and (0.2,-0.2) degrees of correlation, (Cor(u_{1i}, ϵ_i), Cor(u_{2i}, ϵ_i)). Our estimator performs quite similarly to the benchmark OLS estimates that use true underlying treatment status, and both are slightly biased due to endogeneity. This finding is not surprising, considering that our proposed method does not correct for endogeneity.

Third, we allow for non-normal error terms, u_{1i}, u_{2i}, in our bivariate treatment selection model. We consider a number of distributions, including Gamma distribution, bivariate chi-squared distribution, Logistic distribution, Laplace distribution, and exponentially modified Gaussian distribution. The results suggest that the adjusted least squares estimator performs well when the true error terms distribution is symmetrical and poorly otherwise. This tendency may be attributed to the normality assumption that we impose on the distribution of error terms in our proposed approach.

Fourth, we also consider the case where the treatment selection equation is misspecified by including w from a standard normal distribution, x^2, or x^3, in data generating process of t_1 and t_2 but omitting them in the estimation. The proposed estimator is robust to omitting w and x^3 but becomes inconsistent when x^2 is excluded. To minimize bias, we recommend researchers consider whether plausible distribution functions and non-linear model misspecification affect the symmetry of the bivariate error terms distribution in their settings. Lastly, the proposed estimator performs well and remains consistent when we allow the false negative rate in treatment one to be known, while in the second treatment, unknown and estimated using the proposed framework.

1.6 Empirical Example

This section presents our empirical example. Our objective is to compute the adjusted least squares estimates of the effect of SNAP and WIC on food security and HEI. We use the publicly available version of the National Household Food Acquisition and Purchase Survey (FoodAPS). To illustrate the applicability of our methods, we assume other covariates apart from SNAP and WIC are error-free. We also do not account for potential endogeneity in SNAP and WIC participation. We first present naive OLS estimates of the effect of SNAP and WIC on food security. We then provide adjusted least squares estimates using misclassification probabilities of SNAP and WIC available in the literature. Finally, we also ignore the misclassification probabilities available in the literature and we estimate them from the data using our proposed method as a first step and we use them to estimate the treatment effects in the second step.

1.6.1 SNAP and WIC Programs

Households are food secure if they have access to the kinds and quantities of foods needed for an active and healthy life at all times and for all its members, otherwise are food insecure. Food insecurity remains a public health concern in the United States, given the high prevalence among low-income households. Approximately 30.3% of households below 130% of the federal poverty threshold were food insecure in 2021 (Coleman-Jensen et al., 2022a,b). Food insecurity may also lead to poor diet quality, for example, food insecure households may replace fruits, vegetables, and whole grains with calorie-dense and other highly processed foods. In addition, food insecurity is associated with a broad spectrum of detrimental health outcomes (Gundersen et al., 2011; Gundersen and Ziliak, 2018). Inadequate economic resources for food may hinder households from obtaining adequate food or worsen food hardship for already food-insecure households.

SNAP is the largest and WIC the third largest of the 15 domestic food and nutritional assistance programs administered by USDA to address food insecurity and its consequences (Oliveira, 2018). The SNAP program, mean tested, provides nutrition benefits worth more than 60 billion US dollars a year and supplements the food budget of over 42 million individuals on average per month. WIC is also a means-tested federal program that provides nutritional benefits worth more than 5 billion US dollars annually to over 7 million participants on average per month. While SNAP gives vouchers to low-income households to purchase healthy food, WIC provides them vouchers to buy only a restricted set of foods for the nutrition requirements of pregnant, postpartum, and lactating women, infants, and children under five. Moreover, WIC provides counseling and referrals for health services (Bitler et al., 2003). This way, WIC may increase awareness of healthy food choices, and SNAP widens the food choice set.

Misclassification in self-reported SNAP participation status is substantially documented in the literature (e.g., Meyer et al. 2015, Courtemanche et al. 2019, Meyer et al. 2022). Our empirical example uses the misclassification probabilities from Courtemanche et al. (2019). Using FoodAPS data, Courtemanche et al. (2019) provide estimated participation and misclassification error rates for 12 different classification choices of administrative or "gold standard" measures, ADMIN and ALERT, in the survey. We illustrate our empirical application using ADMIN alternate 1, a gold standard measure with the highest misclassification probability, 32.31% false negatives. For WIC, we use misclassification probabilities in Fox and Hokayem (2022). These authors link state administrative data on the WIC program to Current Population Survey Annual Social and Economic Supplement and report a false negative rate of 41.5%. Other empirical evidence about misreporting in WIC can be found in Bitler et al. (2003).

We aim to use the adjusted least squares estimator to account for misreporting in estimating the effect of participating in SNAP and WIC on food insecurity and HEI.

1.6.2 Data

FoodAPS is the first nationally representative survey of US households administered by USDA. The survey is designed to collect comprehensive data about household food purchases, including food obtained through food and nutrition assistance programs. FoodAPS surveys 4,826 households obtained through a multistage sampling design to include low-income households participating in SNAP, low-income households not in SNAP, and higher-income households. In addition to food acquisition, the survey contains self-reported SNAP and WIC participation status, administrative measures of SNAP participation, food security, HEI, income, and demographic characteristics.

We restrict the sample to households simultaneously eligible for SNAP and WIC. These are households with a pregnant woman, a child under five years old, and an income below 130% of the poverty threshold. We use USDA's 30-day food security scores to construct our outcome variable, food insecurity. Our food insecurity score comes from the affirmative responses to the ten items on the household food security questionnaire included in the FoodAPS. The food insecurity scores range from 0 to 10, with 0 representing high food security, and the higher the scores, the greater the degree of food insecurity. We also examine the impact of SNAP and WIC on HEI. Here the HEI index captures the healthfulness or nutritional quality of foods obtained by households by assessing whether they comply with the U.S. Dietary Guidelines for Americans (DGAs). We use the HEI-2010, obtained by summing 12 components and ranges from 0 to 100, where higher scores show greater compliance with the recommended dietary guidelines.

Our empirical example accounts for other control variables. These include dummy variables for gender, education attainment (less than high school, high school or GED, some college, and college degree or higher), ethnicity (Hispanic), race (white, black, or other), employment (employed, searching for a job, or unemployed), and marital status (married, previously married, or never married) of the primary respondent. We also include dummy variables for whether the household lives in the rural census track, whether there are children under five, whether they own any vehicle, and whether the primary store is SNAP-authorized. Our continuous variables include primary respondents' age, household size, household income, number of children, and distance to the primary store in miles. Table 3 present the summary statistics of our outcome variables. The average food insecurity and HEI scores in our sample are 2.7 and 47.5, respectively. In Table 4, we present our summary statistics of other covariates and show the overlap in WIC and SNAP participation. About 42.6% of the sample received both SNAP and WIC, 28.1% received only SNAP, 17.3% WIC only, and 12.1% did not participate in either program.

Table 3. Summary Statistics: Food Insecurity and Healthy Eating

Statistic	Mean	St. Dev.	Min	Max
Food Insecurity Score	2.679	2.665	0	10
Health Eating Index	47.541	11.972	16.310	84.952
Total vegetables	2.525	1.459	0.000	5.000
Greens and beans	1.246	1.696	0.000	5.000
Total fruit	1.984	1.598	0.000	5.000
Whole fruit	2.146	1.871	0.000	5.000
Whole grains	1.650	2.265	0.000	10.000
Dairy	5.934	3.159	0.000	10.000
Total protein foods	3.890	1.380	0.000	5.000
Seafood and plant proteins	1.590	1.701	0.000	5.000
Fatty acids	4.565	3.341	0.000	10.000
Sodium	5.759	3.547	0.000	10.000
Refined grains	5.611	3.503	0.000	10.000
Empty calories	10.642	5.617	0.000	20.000

1.6.3 Results

We present the regression estimates that use given misclassification probabilities in our empirical example in Table 5. The "OLS" columns present estimates from a naive OLS estimator, and the "Adj. LS" columns present results from the adjusted least squares estimators proposed in our study. We use the SNAP false negative rate from Courtemanche et al. (2019), $a_1 = 32.31\%$, and WIC false negative from Fox and Hokayem (2022), $a_1 = 41.5\%$. We present estimates from food insecurity scores in the first two columns. Comparing the results from naive OLS and the adjusted least squares, we observe differences in the magnitude of coefficients across the two estimators. The coefficient of SNAP is almost 42% larger in naive OLS compared to the adjusted least squares. We also observe a plausible expansion bias in the estimated effect of WIC, where the naive OLS estimate coefficient is almost 62% higher than the proposed method estimate. The coefficient of the interaction term is downward biased in the naive OLS, indicating a stronger degree of complementarity than the proposed method. Note that the interaction term estimates are negative in both estimators but 90% lower with naive OLS relative to the proposed method. The last two columns show results using HEI-2010. Here, SNAP and WIC naive OLS coefficient

Table 4. Summary Statistics: Supplemental Nutrition Assistance Program (SNAP), Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) and Demographic Characteristics

Statistic	Mean	St. Dev.	Min	Max
SNAP	0.703	0.458	0	1
WIC	0.599	0.491	0	1
SNAP × WIC	0.425	0.495	0	1
Female	0.889	0.315	0	1
Age (years)	35.056	11.110	16.500	85.000
Less than high school	0.321	0.468	0	1
High school or GED	0.321	0.468	0	1
Some college	0.273	0.446	0	1
College or more	0.085	0.279	0	1
Ethnicity - Hispanic	0.348	0.477	0	1
Race - white	0.611	0.488	0	1
Race - black	0.203	0.403	0	1
Race - others	0.186	0.390	0	1
Employed	0.343	0.475	0	1
Searching for job	0.147	0.355	0	1
Unemployed	0.510	0.501	0	1
Married	0.355	0.479	0	1
Previously married	0.263	0.441	0	1
Never married	0.382	0.486	0	1
Rural	0.232	0.423	0	1
Number of children under 5	1.498	0.793	0	6
Household size	4.746	1.878	1	14
Household income (monthly, 1000$)	1.621	0.922	0.000	5.970
Income to poverty ratio	0.735	0.352	0.000	1.504
Any vehicle indicator	0.749	0.434	0	1
House renting	0.766	0.424	0	1
Food pantry use	0.109	0.312	0	1
Primary store distance (miles)	2.821	4.168	0.078	48.967
Primary store SNAP authorized	0.990	0.098	0	1

estimates are lower, about 89% and 60%, compared to the adjusted least squares estimates, whereas the coefficient for the interaction term is 90% larger relative to the adjusted least squares coefficient estimate. We also compute the average marginal effects and note similarly large differences in magnitude and sign switching in some coefficients. For instance, while SNAP improves healthy eating in our adjusted least squares estimation, the naive OLS estimates suggest that SNAP participation worsens it. Additionally, we provide regression estimates using known misreporting probability and HEI components in Appendix Tables C1 and C2 where for some components, we observe differences in both magnitude and signs in the coefficient estimates of SNAP, WIC and the interaction term across naive OLS and our proposed methods.

We next estimate misreporting probabilities in SNAP and WIC from the data as a first step. Following the method that we proposed in Section 1.4.2, the estimated probabilities of underreporting is $\hat{a}_1 = 29.7\%$ in SNAP and is $\hat{a}_2 = 40.1\%$ in WIC. These estimates are very close to those obtained above from validation studies in the literature. In the second step, we apply the estimated probability in the adjusted least square estimator to obtain the effect of SNAP and WIC on food insecurity and healthy eating. We present the results for food insecurity and HEI in Table 6 and for HEI components in Appendix Tables C3 and C4. Overall, our results and inference are qualitatively similar to when misreporting probabilities are known.

1.7 Conclusion

Literature on measurement errors in binary regressors mainly focuses on single misclassified binary regressors. Our paper exploits the estimation of individual and joint treatment effects in the presence of misclassification errors and other overlapping programs (also plausibly measured with errors). We consider the case of false negatives, which is more common in surveys, and allow the misreporting errors to depend on observable covariates through true binary regressors. We show that the naive OLS estimator is biased, and the joint treatment effect may have a sign opposite to the true effect (sign switching). While complementary programs call for parallel implementation, administering one rather than both may be more cost-effective and welfare-improving if the programs are substitutes. Hence, sign-switching bias may have dramatic

Table 5. Impact of Supplemental Nutrition Assistance Program (SNAP) and Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) on Food Insecurity and Healthy Eating Using Misclassification Probabilities from Validation Data

Dependent Variables:	Food Insecurity Score		Health Eating Index	
	OLS	Adj.LS	OLS	Adj.LS
SNAP	0.211	0.122	-3.102	-0.352
	(0.465)	(0.087)	(2.023)	(0.339)
WIC	0.355	0.135	-1.968	-0.793*
	(0.493)	(0.102)	(2.144)	(0.445)
SNAP × WIC	-0.752	-0.072	4.328*	0.428*
	(0.583)	(0.062)	(2.538)	(0.241)
Female	0.625	0.560	1.458	1.463
	(0.434)	(0.387)	(1.887)	(1.781)
Age	-0.002	-0.002	0.118*	0.124**
	(0.014)	(0.013)	(0.061)	(0.062)
Rural	0.127	0.114	-4.911***	-4.837***
	(0.364)	(0.370)	(1.583)	(1.647)
Household size	0.063	0.060	-0.491	-0.523
	(0.164)	(0.147)	(0.714)	(0.748)
Household income	0.413	0.470	-2.649	-2.737
	(0.596)	(0.533)	(2.595)	(2.441)
Food pantry use	1.101**	1.155**	-0.901	-1.221
	(0.441)	(0.500)	(1.920)	(1.702)

Note: Standard errors in parenthesis and bootstrapped. The analytical sample comes from FoodAPS and consists of households with a pregnant woman or child under five years old and below 130% of the federal poverty threshold. The probability of false negatives in SNAP comes from Courtemanche et al. (2019), $a_1 = 32.31\%$, and in WIC comes from Fox and Hokayem (2022), $a_2 = 41.5\%$. Regressors not reported include primary respondent education, ethnicity, race, employment, marital status, age, rural household residency, indicator and count of children below the age of five, number of children, income to poverty ratio, primary store distance, whether the primary store is SNAP authorized, whether the household has any vehicle and whether the household is renting the house. Significance codes: *** p<0.01, ** p<0.05, * p<0.1

Table 6. Impact of Supplemental Nutrition Assistance Program (SNAP) and Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) on Food Insecurity and Healthy Eating Using Estimated Misclassification Probabilities

Dependent Variables:	Food Insecurity Score		Health Eating Index	
	OLS	Adj.LS	OLS	Adj.LS
SNAP	0.211	0.138	-3.102	-0.392
	(0.465)	(0.099)	(2.023)	(0.387)
WIC	0.355	0.148	-1.968	-0.869*
	(0.493)	(0.111)	(2.144)	(0.485)
SNAP × WIC	-0.752	-0.083	4.328*	0.491*
	(0.583)	(0.071)	(2.538)	(0.278)
Female	0.625	0.557	1.458	1.468
	(0.434)	(0.387)	(1.887)	(1.782)
Age	-0.002	-0.002	0.118*	0.124**
	(0.014)	(0.013)	(0.061)	(0.062)
Rural	0.127	0.115	-4.911***	-4.840***
	(0.364)	(0.370)	(1.583)	(1.648)
Household size	0.063	0.060	-0.491	-0.523
	(0.164)	(0.147)	(0.714)	(0.748)
Household income	0.413	0.472	-2.649	-2.738
	(0.596)	(0.533)	(2.595)	(2.441)
Food pantry use	1.101**	1.154**	-0.901	-1.217
	(0.441)	(0.500)	(1.920)	(1.703)

Note: Standard errors in parenthesis and bootstrapped. The analytical sample comes from FoodAPS and consists of households with a pregnant woman or child under five years old and below 130% of the federal poverty threshold. The probability of false negatives in SNAP is $\hat{a}_1 = 29.7\%$, and in WIC is $\hat{a}_2 = 40.1\%$. The misclassification probabilities are estimated by extending parametric procedure of Hausman et al. (1998) to bivariate models. Regressors not reported include primary respondent education, ethnicity, race, employment, marital status, age, rural household residency, indicator and count of children below the age of five, number of children, income to poverty ratio, primary store distance, whether the primary store is SNAP authorized, whether the household has any vehicle and whether the household is renting the house. Significance codes: *** p<0.01, ** p<0.05, * p<0.1

welfare consequences and lead to suboptimal allocation of resources if used to inform policy, for instance, whether two programs are complements or substitutes and in designing welfare-improving interventions.

Our proposed estimator uses misclassification probabilities to estimate treatment effects consistently. When misclassification probabilities are unknown, we propose a method to estimate them from the data and then apply them in the proposed estimator. It is evident in Monte Carlo simulation that the proposed estimator consistently estimates treatment effects in finite samples and outperforms the naive OLS estimator. Our empirical application, the impact of SNAP and WIC on food insecurity and healthy eating, illustrates differences between the naive OLS and the proposed method in empirical settings. Overall, we document large differences in the magnitude of coefficients across the two estimators and sign switching in the coefficients of the treatment effect model and average marginal effects for some food insecurity and healthy eating outcomes.

Note that our paper does not address the endogeneity in misclassified binary regressors and misreporting. These are potential limitations of our study that are beyond the scope of our current setting, and we leave them to be addressed in future research.

Chapter II: Estimating Treatment Effects with Possibly Missing Participation[1]

2.1 Introduction

Missing data is a common problem in empirical studies that use survey or observational data. For instance, nearly 40% of papers published in the top economics journals, American Economic Review (AER), Journal of Labor Economics (JLE), Journal of Human Resources (JHR), and Quarterly Journal of Economics (QJE), and over half of the empirical papers in JLE and QJE, from 2006 to 2008 have a missing data issue (Abrevaya and Donald, 2017). In addition, Burton and Altman (2004) reviewed journal articles published in 2002 in seven cancer journals and found 81 out of the 100 articles carried out analysis with missing data on covariates. Our paper is motivated by this empirical problem and proposes a method to estimate the effect of a program or policy of interest when participation or treatment status is incomplete (and plausibly endogenous).

To account for missing data in statistical analyses, researchers typically impose assumptions on the plausible mechanism that is driving the missing data. Rubin (1976) introduces three missing data mechanisms: (1) Missing completely at random (MCAR), in which the probability of missing data is unrelated to observable or unobservable data; (2) Missing at random (MAR), where conditional on observables, the probability of missingness is unrelated to unobservable data; and (3) Missing not at random (MNAR), where conditional on observed data, probability of missingness depends on some unobserved data. In this paper, we consider each of these missing data mechanisms in a general framework for estimating treatment effects. In this way, our approach can be extended to missing data in other relevant binary regressors if excluding them from the empirical model could potentially result in omitted variable bias.

Complete case analysis is the most predominant approach in the empirical literature. Abrevaya and Donald (2017) document that most studies, approximately 70%, with missing data issues in top economics journals (AER, JLE, JHR, and QJE) employ the complete case analyses. In complete case analysis, the researcher simply drops observations with missing data, which

[1]This paper is a joint work with Pierre Nguimkeu (Department of Economics, Georgia State University, Email: nnguimkeu@gsu.edu).

makes it easy to implement since it is already a default approach in many statistical software applications. Complete case analysis produces unbiased and consistent estimates when participation is MCAR (see Afifi and Elashoff 1967; Heitjan and Basu 1996; Schafer 1997, among others) and MAR as long as the missing mechanism does not share parameters with the estimated model, which is unlikely if missingness depends on the outcome variable (Rubin, 1976; Heitjan and Basu, 1996; Little and Rubin, 2002; Schafer, 1997; Abrevaya and Donald, 2017). However, researchers may lose valuable information after dropping observations which may substantially reduce the sample size and decrease the efficiency of the estimators, especially when the sample size is small or when a large proportion of data on participation status is missing (e.g.,Afifi and Elashoff 1967; Little 1992; Schafer 1997; Little and Rubin 2002; Abrevaya and Donald 2017; Dardanoni et al. 2011). In addition, in many empirical applications, treatment status is MNAR. For example, in smoking literature, there is a plausible association between smoking status and the probability that this status is observed in the data (Jackson et al., 2014).

Other studies replace missing data with approximations or imputed values. This method is also common in empirical studies. As documented in Abrevaya and Donald (2017), approximately 20% of studies published in top economic journals (AER, JLE, JHR, and QJE) use imputation methods. The most common data imputation techniques include unconditional mean imputation, conditional mean imputation, and multiple imputations (e.g., Afifi and Elashoff 1967, Dagenais 1973; Rubin 1988; Little 1992; Little and Rubin 2002; Rubin 2004; Horton and Kleinman 2007; Abrevaya and Donald 2017; McDonough and Millimet 2017; Galimard et al. 2018). In addition, for binary regressor, Subasi et al. (2011) proposes an imputation method based on a measure of similarity.

Naive approaches to impute missing values are typically not recommended (Little, 1992; Little and Rubin, 2002). In addition, measurement errors in imputed values may lead to biased and inconsistent estimates (Horton and Kleinman, 2007; Dardanoni et al., 2011), and, even with multiple imputations, stronger assumptions may be required to rule out these measurement errors (Greene, 2012). Imputation in binary variables is also not straightforward as it may be affected by a perfect prediction, which may bias treatment effect estimates (White et al., 2010), and may require researchers to specify a method of converting the imputed values into binary values (Bernaards et al., 2007). Moreover, the efficacy of imputation methods when missingness occurs in binary regressors is not well developed (Greene, 2012).

Other studies use the missing indicator approach, which replaces the missing data with zero and includes an additional binary variable that indicates whether the data value is missing. The missing indicator method, used by approximately 20% of papers reviewed by Abrevaya and Donald (2017), is generally considered inconsistent even when data is MCAR unless the missing covariate is orthogonal to other covariates in the model (Jones, 1996; Horton and Kleinman, 2007; Abrevaya and Donald, 2017). Other approaches employed in empirical literature include maximum likelihood estimation and Bayesian methods (e.g., Maity et al. 2019; Galimard et al. 2018).

Our contribution to the literature is in three-fold. First, we propose a parametric regression framework for estimating treament effect when participation is plausibly endogenous and MNAR, for which MAR and MCAR are special cases. Second, we present and discuss theoretical asymptotic bias in Ordinary Least Squares (OLS) and Instrumental Variables (IV) estimators when complete case analysis is used. In particular, we show that OLS and IV estimators are biased and inconsistent when data is not MCAR as long as missingness is correlated with the outcome variable (in addition to the bias that affects the OLS estimator when participation is endogenous). We demonstrate that the bias in complete case analyses can go either way and, in some settings, complete case analysis returns estimates that have an opposite sign to the true values, a problem commonly referred to as sign switching or sign reversal. Third, we propose a method to consistently estimate the parameters of interest in our treatment effect framework, including the impact of a program or policy of interest that has missing data on participation status. Our approach proceeds in three steps. The first step assigns all missing data to non-participation status. Steps two and three correct for misclassification using a method proposed in Nguimkeu et al. (2019). Particularly, step two employs a partial observability model to estimate true participation, and step three estimates the treatment effect of interest using the predicted probability of true participation status obtained from step two. Our Monte Carlo simulations show that the proposed estimator performs well in finite samples and outperforms the comparative estimators, including complete case analysis OLS and IV estimators and the missing-indicator method, when data is not MCAR. Finally, we provide an empirical illustration, estimating the effect of maternal smoking on birth weight using the 1989 U.S. Natality data.

The intuition behind our three-step estimator is simple. In the first step, all program non-participants are correctly allocated, while participants are misclassified. Consequentially, this step transforms our missing data problem into that of one-sided endogenous misreporting, allowing us to leverage the extensive literature on measurement errors in binary regressors. Although a number of solutions have been developed to address measurement error in binary regressors (for instance, see, Aigner 1973, Bollinger (1996), Kane et al. 1999, Black et al. 2000, Frazis and Loewenstein 2003, Lewbel 2007, Chen et al. 2008b, Chen et al. 2008a, Brachet 2008, Kreider 2010, Kreider (2010), van Hasselt and Bollinger 2012, DiTraglia and García-Jimeno 2017, Ura 2018, Nguimkeu et al. 2019, Nguimkeu et al. 2021, among others), we focus on Nguimkeu et al. (2019) as their framework accounts for both endogenous one-sided misreporting and endogenous participation. Identification in our proposed framework relies on the availability of an instrumental variable for participation that is not correlated with missingness and an additional random variable that is associated with missingness but does not need to be uncorrelated with the outcome. Hence, the misclassified treatment status from the first step can be modeled as a joint function of true treatment status and a missing indicator, which makes it possible to estimate the probability of true participation status in the second step and identify point estimates of the treatment effect in the third step.

The rest of the paper is structured as follows. Section 2.2 presents the missing program participation status model and shows that OLS and IV complete case analysis leads to inconsistent estimates. Section 2.3 describes the proposed estimator. Section 2.4 presents Monte

Carlo simulations, section 2.5 gives an empirical example, and section 2.6 concludes. Mathematical proofs are in the Appendix D.

2.2 The Model

This section presents our framework, describes our estimation strategy, and discusses asymptotic bias in complete case analysis.

2.2.1 The Set-Up

Consider a standard linear regression model where for a random sample of size n, a researcher observes the outcome y_i, the vectors of covariates x_i, z_i and w_i of sizes $k \times 1$, $p \times 1$ and $q \times 1$, respectively, for all observations in the data and the indicator for participation, T_i, which is incomplete. Thus, for many individuals in data, the value of T_i is unknown.

The relationship between these entities is the treatment effect model defined by

$$y_i = \mathrm{x}_i'\gamma + T_i\beta + \varepsilon_i, \tag{12}$$

where the binary treatment, T_i is determined by a vector of exogenous covariates, z_i and reporting indicator, R_i such that

$$T_i = \begin{cases} 1\left[\mathrm{z}_i'\delta + u_i > 0\right] & \text{if} \quad R_i = 1 \\ \\ \text{Not available} & \text{if} \quad R_i = 0 \end{cases} \tag{13}$$

and the error term u_i is correlated with ε_i with correlation coeffcient φ_u. When $\varphi_u = 0$, then T_i and ε_i are uncorrelated, and participation is said to be exogenous. However, when $\varphi_u \neq 0$, then $\mathbb{E}[\varepsilon_i|T_i] \neq 0$ and participation is endogenous. For the reporting indicator R_i, we consider a general missing pattern of the binary treatment that follows a missing mechanism such that

$$R_i = 1\left[\mathrm{w}_i'\theta + v_i > 0\right] \tag{14}$$

where the error term v_i is correlated with ε_i with correlation coefficient φ_v. The missing mechanism is, therefore, similar to Heckman sample selection driven by the observable covariates w_i.

Throughout the paper, the joint distribution of the error terms (u_i, v_i) is assumed to follow a bivariate normal distribution conditionally on $(\mathrm{x}_i, \mathrm{z}_i, \mathrm{w}_i)$.

Assumption 5. *Conditional on covariates* $\mathrm{x}_i, \mathrm{z}_i, \mathrm{w}_i$, *the error terms* (u_i, v_i) *follow bivariate normal distribution, given by*

$$\begin{pmatrix} u_i \\ v_i \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right). \tag{15}$$

where $\rho \in [0, 1)$ corresponds to the correlation coefficient between the error terms of the participation equation and the reporting equation. Notice that when $\rho = 0$, then u_i and v_i are independent and $\Pr[R_i = 1|T_i, \mathrm{z}_i, \mathrm{w}_i]$ does not depend on T_i. This corresponds to the missing mechanism known as MAR. However, when $\rho \neq 0$, then u_i and v_i are dependent and $\Pr[R_i = 1|T_i, \mathrm{z}_i, \mathrm{w}_i]$ depends on T_i. The resulting missing mechanism is known as MNAR. Moreover, increasing values of ρ correspond to increasing strength of the MNAR mechanism. Finally, when $\rho = 0$, $\varphi_v = 0$, and $\theta = 0$, then $\Pr[R_i = 1|T_i, \mathrm{z}_i, \mathrm{w}_i]$ is a constant that does not depend on any covariates and the missing mechanism is known as MCAR.

Assumption 6. $\mathbb{E}[\varepsilon_i|\mathrm{x}_i, \mathrm{z}_i] = 0$ *and* $\mathbb{E}[\varepsilon_i^2|\mathrm{x}_i, \mathrm{z}_i] = \sigma^2$,

In particular, the error term, ε_i, is independent of exogenous covariates in the outcome model, $\mathrm{x_i}$, and in the binary treatment model, $\mathrm{z_i}$

Assumption 7. *The* $k \times k$ *matrix,* $\mathbb{E}[\mathrm{x}_i\mathrm{x}_i']$ *exists and is nonsingular.*

The vector of model parameters is $\psi = (\beta, \gamma', \delta', \theta', \rho, \varphi_u, \varphi_v, \sigma^2)'$, where β is the main parameter of interest and represents the effect of the treatment T on the outcome y that we wish to estimate.

2.2.2 The Complete-Case Analysis

The standard treatment of missing data in practice is the complete-case analysis, sometimes referred to as listwise deletion, where cases with any missing values are simply discarded. The main advantage is the ease of implementation. The complete-case analysis, therefore, uses only complete observations to estimate the treatment effect.

Denote by $R = \begin{pmatrix} R_1 & 0 & \dots & 0 \\ 0 & R_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & R_n \end{pmatrix}$ the (symmetric and idempotent) matrix of reporting indicators (i.e., R_i is one if the ith individual has complete data for T and is zero otherwise).

Then the model for complete data is defined by

$$Ry = RX\gamma + RT\beta + R\varepsilon \tag{16}$$

that is

$$\tilde{y} = \tilde{X}\gamma + \tilde{T}\beta + R\varepsilon \tag{17}$$

where for a any variable H, we have denoted $\tilde{H} = RH$

We consider both the naive ordinary least squares (OLS) estimator and the instrumental variable (IV) estimator derived from the complete case analysis. The OLS estimator of β is obtained as

$$\widehat{\beta}_{LS} = \frac{\tilde{T}\tilde{M}\tilde{y}}{\tilde{T}\tilde{M}\tilde{T}} = \frac{\sum_{i=1}^{n} T_i R_i y_i - \sum_{i=1}^{n} T_i R_i \mathrm{x}_i \left(\sum_{i=1}^{n} \mathrm{x}_i \mathrm{x}_i' R_i\right)^{-1} \sum_{i=1}^{n} y_i R_i \mathrm{x}_i}{\sum_{i=1}^{n} T_i R_i - \sum_{i=1}^{n} T_i R_i \mathrm{x}_i \left(\sum_{i=1}^{n} \mathrm{x}_i \mathrm{x}_i' R_i\right)^{-1} \sum_{i=1}^{n} T_i R_i \mathrm{x}_i} \tag{18}$$

where $\tilde{M} = I - \tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}' = I - RX\,(X'RX)^{-1}\,X'R$.

The complete-case IV estimator is a two-stage least squares estimator where the first stage is a probit for participation using complete data, $\Pr[\tilde{T}_i = 1 | \mathrm{z}_i, R_i] = \Phi(\mathrm{z}_i\delta)$, and the predicted values of this probit, $\tilde{\Phi}_i = \Phi(z_i \hat{\boldsymbol{\delta}}_c)$, are used in lieu of $\tilde{T}_i$ in the outcome equation (17). The IV estimator of β derived as the OLS of this second-stage regression is given by

$$\begin{aligned}\widehat{\beta}_{IV} &= \frac{\tilde{\Phi}\tilde{M}\tilde{y}}{\tilde{\Phi}\tilde{M}\tilde{\Phi}} \\ &= \frac{\sum_{i=1}^{n} \Phi(z_i R_i \hat{\boldsymbol{\delta}}_c) R_i y_i - \sum_{i=1}^{n} \Phi(z_i R_i \hat{\boldsymbol{\delta}}_c) R_i x_i \left(\sum_{i=1}^{n} x_i x_i' R_i\right)^{-1} \sum_{i=1}^{n} y_i R_i x_i}{\sum_{i=1}^{n} \Phi(z_i R_i \hat{\boldsymbol{\delta}}_c)^2 R_i - \sum_{i=1}^{n} \Phi(z_i R_i \hat{\boldsymbol{\delta}}_c) R_i x_i \left(\sum_{i=1}^{n} x_i x_i' R_i\right)^{-1} \sum_{i=1}^{n} \Phi(z_i R_i \hat{\boldsymbol{\delta}}_c) R_i x_i}\end{aligned} \tag{19}$$

where $\hat{\delta}_c$ is the MLE estimator of δ under the complete-case scenario.

To understand the nature of the asymptotic biases associated with these estimators, we assume the normality of the error terms in the outcome equation, which allows us to obtain functional forms.

Assumption 8. $\varepsilon_i | x_i, z_i, w_i \sim \mathcal{N}(0, \sigma^2)$.

We then have the following results.

Theorem 2. *Under the model assumptions 5-8, the probability limits of the OLS and IV estimators are given by*

$$\beta_{LS} \xrightarrow{p} \beta + \sigma\left[\varphi_u A(\rho) + \varphi_v B(\rho)\right] \tag{20}$$

and

$$\beta_{IV} \xrightarrow{p} \beta + \sigma\varphi_v C \tag{21}$$

where the quantities $A(\rho)$, $B(\rho)$, and C are given by:

$$A(\rho) = \frac{\mathbb{E}\left[\phi\left(z_i'\delta\right)\Phi\left(\frac{w_i'\theta - \rho z_i'\delta}{\sqrt{1-\rho^2}}\right)\right]}{\mathbb{E}(T_i R_i) - \mathbb{E}(T_i R_i x_i)\mathbb{E}(x_i x_i' R_i)^{-1}\mathbb{E}[T_i R_i x_i']},$$

$$B(\rho) = \frac{\mathbb{E}\left[\phi\left(\mathrm{w}_i'\theta\right)\Phi\left(\frac{\mathrm{z}_i'\delta - \rho \mathrm{w}_i'\theta}{\sqrt{1-\rho^2}}\right)\right] - \mathbb{E}(T_i R_i \mathrm{x}_i')\mathbb{E}(\mathrm{x}_i \mathrm{x}_i' R_i)^{-1}\mathbb{E}[\mathrm{x}_i \phi\left(\mathrm{w}_i'\theta\right)]}{\mathbb{E}(T_i R_i) - \mathbb{E}(T_i R_i \mathrm{x}_i)\mathbb{E}(\mathrm{x}_i \mathrm{x}_i' R_i)^{-1}\mathbb{E}[T_i R_i \mathrm{x}_i']}$$

$$C = \frac{\mathbb{E}\left[\Phi(\mathrm{z}_i'\delta_c)\phi\left(\mathrm{w}_i'\theta\right)\right] - \mathbb{E}\left[\Phi(\mathrm{z}_i'\delta_c)\Phi\left(\mathrm{w}_i'\theta\right)\mathrm{x}_i\right]\mathbb{E}\left[\mathrm{x}_i \mathrm{x}_i' R_i\right]^{-1}\mathbb{E}\left[\mathrm{x}_i \phi\left(\mathrm{w}_i'\theta\right)\right]}{\mathbb{E}\left[\Phi(\mathrm{z}_i'\delta_c)^2\Phi\left(\mathrm{w}_i'\theta\right)\right] - \mathbb{E}\left[\Phi(\mathrm{z}_i'\delta_c)\Phi\left(\mathrm{w}_i'\theta\right)\mathrm{x}_i'\right]\mathbb{E}\left[\mathrm{x}_i \mathrm{x}_i' R_i\right]^{-1}\mathbb{E}\left[\Phi(\mathrm{z}_i'\delta_c)\Phi\left(\mathrm{w}_i'\theta\right)\mathrm{x}_i\right]}$$

where δ_c is the probability limit of $\hat{\delta}_c$.

Proof. See Appendix D.1.

Bias in the complete case analysis depends on the mechanism that seems to be driving the missing data. With exogenous treatment, $\varphi_u = 0$, and MCAR, both OLS and IV are unbiased. If treatment is endogenous but MCAR, then the IV estimator is consistent, while the OLS estimator is inconsistent, driven by the endogeneity of treatment assignment. If participation is exogenous, but MAR or MNAR, then OLS and IV estimators are inconsistent as long as the missing mechanism is correlated with the outcome of interest, $\varphi_v \neq 0$. This generalization extends to MAR and MNAR mechanisms with endogenous treatment. However, in this case, non-random treatment assignment will also add to the bias of the OLS estimator.

Theorem 2 suggests that the direction and magnitude of the bias in the OLS estimator are determined by $A(\rho)$, $B(\rho)$, ρ, φ_v, and φ_u while in the IV estimator by C and φ_v. Depending on these values, complete case OLS and IV estimators can suffer from either attenuation or expansion bias. For instance, if $\beta > 0$, $A(\rho) > 0$, and $B(\rho) > 0$, complete case OLS estimates will underestimate the treatment effects if probabilities of missing data and treatment are negatively correlated with the outcome of interest, thus $\varphi_v < 0$ and $\varphi_u < 0$, and overestimate treatment effect if reporting and treatment are positively correlated with the outcome, $\varphi_v > 0$ and $\varphi_u > 0$. Additionally, when $\varphi_v < 0$ and $\varphi_u > 0$ (or $\varphi_v > 0$ and $\varphi_u < 0$), OLS estimates will be downward bias when $|\varphi_v B(\rho)| > |\varphi_u A(\rho)|$ (or $|\varphi_v B(\rho)| < |\varphi_u A(\rho)|$), and upward biased if

$|\varphi_v B(\rho)| < |\varphi_u A(\rho)|$ (or $|\varphi_v B(\rho)| > |\varphi_u A(\rho)|$). Similarly, if $C > 0$, we have attenuation bias in the IV estimator when $\varphi_v < 0$ and expansion bias when $\varphi_v > 0$.

The treatment effect estimates in complete case analysis may also have opposite signs to the true treatment effect. For example, when $\beta > 0$, β_{IV} will switch sign when $0 < \beta < -\sigma\varphi_v C$ and the sign of β_{LS} will be reversed when $0 < \beta < -\sigma\,[\varphi_u A(\rho) + \varphi_v B(\rho)]$. Sign-switching may lead to erroneous conclusions regarding whether or not the treatment is improving the outcome, which can have adverse repercussions if used to inform policy decisions.

2.3 The Proposed Three-Steps Estimator

Our primary interest is to consistently estimate the treatment effect, β in the specification (12), where the participation status, T_i, is plausibly endogenous and missing for some observations in our data. The estimation approach proposed in this paper frames the issue of missing data in treatment status as a partial observability problem in a similar spirit to Poirier (1980) and Nguimkeu et al. (2019). Our proposed estimation approach proceeds in three steps:

First step: For all n observations, construct a partially observable variable $\check{T}_i$, such that $\check{T}_i = T_i R_i$, where $T_i = \mathrm{z}_i'\delta + u_i$ and $R_i = \mathrm{w}_i'\theta + v_i$.

Second step: Estimate parameters δ and θ by using $\check{T}_i$ by employing a partial observability probit model, following joint normal distribution of u_i and v_i described in Assumption (5) and compute $\hat{T}_i = \phi(\mathrm{z}_i'\hat{\delta})$, the predicted probability of individual i treatment status.

Third step: Estimate the treatment effect, β, in the model given by Equation 12 using the predicted treatment status, $\hat{T}_i$, obtained in the second step in lieu of the observed status with missing data, T_i.

Our first step involves the construction of a partially observable variable, $\check{T}_i$, for all observations in the data. Note that this step simply translates to replacing the missing data with zeros as follows:

$$\check{T}_i = \begin{cases} T_i & \text{if} \quad R_i = 1 \\ 0 & \text{if} \quad R_i = 0 \end{cases}$$

Consequentially, all treated observations are correctly allocated, while some participants are misclassified. This presence of false negatives in the newly constructed participation status turns our missing data problem into that of plausibly endogenous one-sided misreporting in binary regressors, whose solution exists from Nguimkeu et al. (2019).

Following Poirier (1980) and Nguimkeu et al. (2019), the binary choice model in the second step can be expressed as

$$\Pr(\check{T}_i = 1|\mathrm{z}_i, \mathrm{w}_i) = \Pr(-u_i > \mathrm{z}_i'\delta, -v_i > \mathrm{w}_i'\theta) = \mathrm{F}_{u,v}(\mathrm{z}_i'\delta, \mathrm{w}_i'\theta, \rho) = \mathrm{P}_i(\delta, \theta, \rho)$$

From joint normal distribution of u_i and v_i , parameters of the binary choice model, δ, θ, and ρ can be estimated jointly by maximizing the following likelihood function

$$\mathcal{L}(\delta, \theta, \rho) = \frac{1}{n} \sum_{i=1}^{n} \{\check{T}_i \ln P_i(\delta, \theta, \rho) + (1 - \check{T}i) \ln P_i(\delta, \theta, \rho)\}$$

Under correct model distribution assumptions, the maximum likelihood estimates of the model parameter, including $\hat{\delta}$, are consistent. Using $\hat{\delta}$, the probability of true treatment status can be computed using $\hat{T}_i = \phi(\mathrm{z}_i'\hat{\delta})$.

The third step estimates the treatment effect, β, using the predicted treatment status, $\hat{T}_i$, obtained from the second step. Specifically, we estimate

$$y_i = x_i'\gamma + \hat{T}_i\beta + \zeta_i$$

The proposed treatment effect estimator follows as

$$\hat{\beta}_{3s} = (\hat{T}'M_x\hat{T})^{-1}\hat{T}'M_xy$$

Where $M_x = I - X(X'X)^{-1}X'$

Theorem 3. *Assumptions 5-7 hold. The three-step estimator of treatment effect is consistent, i.e* $\hat{\beta}_{3s} \xrightarrow{p} \beta$

Proof. See Appendix D.2.

2.4 Monte Carlo Simulations

In this section, we assess the performance of the proposed estimator in finite samples through Monte Carlo simulations and compare our estimator with other approaches, including the complete-case analysis OLS estimator, complete-case analysis IV estimator, and the missing indicator method. Our goal is to consistently estimate, β, the conditional average treatment effect of T, on our outcome of interest, y, given by Equation (12), assuming that treatment status, T_i, is missing for some observations while the control variables, x_i, are observed for all n observations. In addition, as discussed below, our Monte Carlo simulations also examine the robustness of the proposed estimator to model misspecifications and slight deviations from the assumptions we previously stated.

2.4.1 Simulation Setup

Our data generation process proceeds as follows. The treatment indicator, T_i, is given by

$$T_i = 1\left[\delta_0 + \mathrm{z}_i'\delta_1 + u_i > 0\right]$$

Where $\delta_0 = 0.1$, $\delta_1 = [1 \;\; 1]$, $\mathrm{z}_i = [\mathrm{z}_{1i} \;\; \mathrm{z}_{2i}]$, $\mathrm{z}_{1i} \sim \mathcal{N}(0,1)$, and $\mathrm{z}_{2i} \sim \mathcal{N}(0,1)$.

The treatment effect model is given by

$$y_i = \gamma_0 + x_i\gamma_1 + T_i\beta + \varepsilon_i,$$

where $\gamma_0 = 1$, $\gamma_1 = 1$, $\beta = 0.3$, and $\mathrm{x}_i \sim \mathcal{N}(0,1)$. We aim to estimate the model parameters consistently, and of most interest is the treatment effect estimate, $\hat{\beta}$.

The econometrician observes the treatment indicator, T_i, for a fraction of observations determined by a threshold, c. The following reporting indicator determines whether the treatment status of observation i is observed in the data and the mechanism driving the missing data,

$$R_i = 1\left[\theta_0 + \mathrm{w}_i'\theta_1 + v_i > c\right]$$

where $\theta_0 = 0.1$, $\theta_1 = [1 \;\; 1]$, and $\mathrm{w}_i = [\mathrm{w}_{1i} \;\; \mathrm{w}_{2i}]$, such that $\mathrm{w}_{1i} \sim \mathcal{N}(0,1)$ and $\mathrm{w}_{2i} = \mathrm{z}_{2i}$. The error terms, ε_i, u_i, and v_i come from a trivariate normal distribution defined by

$$\begin{bmatrix} \epsilon_i \\ u_i \\ v_i \end{bmatrix} \sim \mathcal{N}\left[\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma & \sigma\varphi_u & \sigma\varphi_v \\ \sigma\varphi_u & 1 & \rho \\ \sigma\varphi_v & \rho & 1 \end{bmatrix}\right]$$

where $\sigma = 1$. We vary the values of φ_u to explore different degrees of endogeneity of the treatment status. In addition, we also carry out our simulations using different values of φ_v and ρ to assess the performance of our estimator under different levels of correlations of the reporting indicator with the treatment status and the outcome of interest. This setup also allows us to exploit the performance of our approach under different missing data mechanisms, such that participation status is MAR when $\varphi_v \neq 0$ and $\rho = 0$, and MNAR if $\rho \neq 0$. Specifically, we let $\varphi_u \in \{0, 0.4\}$, $\varphi_v \in \{0, 0.3, 0.7\}$, and $\rho \in \{0, 0.3\}$. In all these scenarios, we choose the threshold c such that 0%, 5%, 10%, 20%, and 30% of the participation status is missing to examine the bias discussed in Section 2.2 and performance of our proposed estimator across different severity of the missing data issue.

We first employ an IV estimator using the entire dataset to estimate the model parameters, assuming the participation status is fully observed. These estimates provide a baseline scenario for comparison with estimates from the OLS estimator, complete-case analysis, and our proposed solution. In a complete case analysis, we estimate the treatment effects using the OLS estimator, $\widehat{\beta}_{LS}$ and IV estimator, $\widehat{\beta}_{IV}$. We also consider other estimators that are used in the empirical literature. In particular, we employ a variant of the missing indicator method known as modified zero-order regression. We obtain the treatment effect estimates in this missing-indicator approach by setting the missing participation status, T_i, to zero and estimate the treatment effect model that includes the reporting indicator, R_i, and the interaction between the reporting and treatment indicators. Finally, we use our proposed method, the three-step estimator, to estimate the effect of T_i, $\widehat{\beta}_{3s}$.

2.4.2 Simulation Results

We execute 1000 replications using the sample size of n=5000 observations and report the average of the simulation results in Table 7. In the first two columns, "IV No Missing" and "OLS No Missing," we present IV and OLS estimates when the full data is used in estimation, assuming that treatment status is observed for all units. "IV No Missing" provides a benchmark for comparing "OLS No Missing," complete case analysis, the missing indicator method, and our proposed estimator. "IV CC" and "OLS CC" show IV and OLS estimates obtained after dropping out observations with missing data on participation status, thus complete-case analysis. We present the results of the modified zero-order regression in the "Missing Indicator" column, and the estimates from our proposed estimator are presented in the last column, "Three-Step Estimator."

The simulation results in Table 7 align well with our theoretical discussions. As expected, the IV estimator is consistent when we utilize all observations in the simulated data, assuming that treatment status is fully observed. The OLS estimator is consistent only when the participation is exogenous. In complete-case analysis, the IV estimator is downward biased and inconsistent when the treatment is MAR and MNAR, and the bias worsens as the ratio of missing data increases. As shown in Theorem 2, the complete case OLS estimator is consistent when the treatment indicator is MCAR and exogenous. Otherwise, we observe an attenuation bias when participation is MAR or MNAR and exogenous and an expansion bias when the treatment is MAR or MNAR and endogenous. The missing indicator method performs similarly to the complete-case analysis OLS estimator. Turning to our proposed estimator in the last column, the three-step estimator is unbiased when treatment is MCAR and remains unbiased in MAR and MNAR cases as proposed in Theorem 3. The proposed method is also consistent when treatment

Table 7. Monte Carlo Simulation Results

%Missing	φ_v	φ_u	ρ	IV No Missing	OLS No Missing	IV CC	OLS CC	Missing Indicator	Three-Step Estimator
0%	0	0	0	0.3006	0.3008	0.3006	0.3008		0.3005
			0.3	0.3006	0.3008	0.3006	0.3008		0.3005
		0.4	0	0.3003	0.6694	0.3003	0.6694		0.3002
			0.3	0.3003	0.6694	0.3003	0.6694		0.3002
	0.3	0	0	0.3005	0.3006	0.3005	0.3006		0.3004
			0.3	0.3005	0.3006	0.3005	0.3006		0.3004
		0.4	0	0.3002	0.6692	0.3002	0.6692		0.3001
			0.3	0.3002	0.6692	0.3002	0.6692		0.3002
	0.7	0	0	0.3003	0.3002	0.3003	0.3002		0.3002
			0.3	0.3003	0.3001	0.3003	0.3001		0.3002
		0.4	0	0.3000	0.6687	0.3000	0.6687		0.2999
			0.3	0.3000	0.6689	0.3000	0.6689		0.3000
5%	0	0	0	0.3006	0.3008	0.2995	0.3004	0.3004	0.3008
			0.3	0.3006	0.3008	0.2999	0.3004	0.3004	0.3008
		0.4	0	0.3003	0.6694	0.2992	0.6733	0.6733	0.3005
			0.3	0.3003	0.6694	0.2996	0.6761	0.6761	0.3006
	0.3	0	0	0.3005	0.3006	0.2607	0.2826	0.2826	0.3007
			0.3	0.3005	0.3006	0.2612	0.2770	0.2770	0.3008
		0.4	0	0.3002	0.6692	0.2604	0.6555	0.6554	0.3004
			0.3	0.3002	0.6692	0.2609	0.6528	0.6528	0.3005
	0.7	0	0	0.3003	0.3002	0.2091	0.2587	0.2587	0.3005
			0.3	0.3003	0.3001	0.2095	0.2458	0.2458	0.3005
		0.4	0	0.3000	0.6687	0.2088	0.6315	0.6315	0.3002
			0.3	0.3000	0.6689	0.2092	0.6216	0.6216	0.3003
10%	0.0	0	0	0.3006	0.3008	0.2999	0.3007	0.3007	0.3008
			0.3	0.3006	0.3008	0.3000	0.3007	0.3007	0.3009
		0.4	0	0.3003	0.6694	0.2994	0.6760	0.6760	0.3005
			0.3	0.3003	0.6694	0.2994	0.6808	0.6808	0.3006
	0.3	0	0	0.3005	0.3006	0.2400	0.2734	0.2734	0.3007
			0.3	0.3005	0.3006	0.2405	0.2645	0.2645	0.3008
		0.4	0	0.3002	0.6692	0.2395	0.6487	0.6487	0.3004
			0.3	0.3002	0.6692	0.2399	0.6446	0.6446	0.3005
	0.7	0	0	0.3003	0.3002	0.1601	0.2368	0.2368	0.3005
			0.3	0.3003	0.3001	0.1611	0.2161	0.2161	0.3005
		0.4	0	0.3000	0.6687	0.1596	0.6121	0.6121	0.3002
			0.3	0.3000	0.6689	0.1606	0.5962	0.5962	0.3003
20%	0	0	0	0.3006	0.3008	0.2998	0.3004	0.3004	0.3008
			0.3	0.3006	0.3008	0.3001	0.3007	0.3007	0.3008
		0.4	0	0.3003	0.6694	0.2989	0.6794	0.6794	0.3006
			0.3	0.3003	0.6694	0.2995	0.6875	0.6875	0.3006
	0.3	0	0	0.3005	0.3006	0.2101	0.2599	0.2599	0.3007
			0.3	0.3005	0.3006	0.2114	0.2463	0.2463	0.3007
		0.4	0	0.3002	0.6692	0.2092	0.6389	0.6389	0.3005
			0.3	0.3002	0.6692	0.2108	0.6332	0.6332	0.3005
	0.7	0	0	0.3003	0.3002	0.0905	0.2059	0.2059	0.3006
			0.3	0.3003	0.3001	0.0931	0.1737	0.1737	0.3005
		0.4	0	0.3000	0.6687	0.0896	0.5849	0.5849	0.3003
			0.3	0.3000	0.6689	0.0925	0.5607	0.5607	0.3003
30%	0.0	0	0	0.3006	0.3008	0.2999	0.3006	0.3005	0.3007
			0.3	0.3006	0.3008	0.3003	0.3006	0.3006	0.3008
		0.4	0	0.3003	0.6694	0.2990	0.6827	0.6827	0.3005
			0.3	0.3003	0.6694	0.2996	0.6933	0.6933	0.3006
	0.3	0	0	0.3005	0.3006	0.1880	0.2506	0.2506	0.3007
			0.3	0.3005	0.3006	0.1901	0.2329	0.2329	0.3007
		0.4	0	0.3002	0.6692	0.1871	0.6327	0.6328	0.3005
			0.3	0.3002	0.6692	0.1894	0.6256	0.6256	0.3005
	0.7	0	0	0.3003	0.3002	0.0389	0.1839	0.1839	0.3005
			0.3	0.3003	0.3001	0.0431	0.1425	0.1425	0.3005
		0.4	0	0.3000	0.6687	0.0379	0.5660	0.5660	0.3003
			0.3	0.3000	0.6689	0.0424	0.5352	0.5352	0.3004

is plausibly endogenous and remains unbiased with increased severity of the missing data problem.

We explore the sensitivity of our three-step estimator to changes in the main assumptions of the proposed framework and misspecification in the operational model that the researcher estimates. In particular, Table 8 presents Monte Carlo simulation results with non-normal errors where, in separate settings, we generate ϵ_i, u_i, and v_i from trivariate chri-square distribution, gamma distribution, and exponential distribution. Overall, our estimator remains robust and performs well relative to our baseline scenario, IV estimation, assuming that participation is completely observed. We introduce misspecification in the partial observability model and present the corresponding simulation results in Table 9. In columns 5 and 6, we assume the researcher does not observe the additional variable that is correlated with missingness and estimates the probability of true participation status in the second step using only covariates in the treatment effect model and the instrumental variable for participation. Results in columns 7 to 10 are obtained when additional variables, w^2 or w^3, influence the missing data mechanism in the data generating process but are omitted in the model estimation. The proposed estimator performs quite well in these settings and remains robust across different missing data mechanisms and degrees of missing data.

Table 8. Monte Carlo Simulation Results: Robustness to the Distribution of Error Assumptions

%Missing	φ_v	φ_u	ρ	IV No Missing χ_1^2	Three-Step Estimator χ_1^2	IV No Missing $\Gamma(2,1)$	Three-Step Estimator $\Gamma(2,1)$	IV No Missing Exponential (1)	Three-Step Estimator Exponential (1)
0%	0	0	0	0.3075	0.3076	0.2960	0.2962	0.3044	0.3044
			0.3	0.3075	0.3076	0.2960	0.2962	0.3044	0.3044
		0.4	0	0.3070	0.3071	0.2951	0.2953	0.3040	0.3041
			0.3	0.3072	0.3073	0.2950	0.2953	0.3041	0.3042
	0.3	0	0	0.3070	0.3071	0.2961	0.2964	0.3041	0.3042
			0.3	0.3070	0.3070	0.2961	0.2964	0.3041	0.3041
		0.4	0	0.3061	0.3062	0.2949	0.2952	0.3036	0.3036
			0.3	0.3065	0.3066	0.2950	0.2953	0.3038	0.3038
	0.7	0	0	0.3054	0.3055	0.2956	0.2959	0.3033	0.3033
			0.3	0.3053	0.3053	0.2956	0.2958	0.3032	0.3032
		0.4	0	0.3039	0.3039	0.2939	0.2941	0.3023	0.3024
			0.3	0.3045	0.3046	0.2942	0.2944	0.3027	0.3027
5%	0	0	0	0.3075	0.3044	0.2960	0.2919	0.3044	0.3033
			0.3	0.3075	0.3057	0.2960	0.2945	0.3044	0.3041
		0.4	0	0.3070	0.3039	0.2951	0.2911	0.3040	0.3030
			0.3	0.3072	0.3054	0.2950	0.2936	0.3041	0.3038
	0.3	0	0	0.3070	0.3038	0.2961	0.2921	0.3041	0.3030
			0.3	0.3070	0.3051	0.2961	0.2946	0.3041	0.3038
		0.4	0	0.3061	0.3030	0.2949	0.2909	0.3036	0.3025
			0.3	0.3065	0.3047	0.2950	0.2936	0.3038	0.3035
	0.7	0	0	0.3054	0.3022	0.2956	0.2916	0.3033	0.3022
			0.3	0.3053	0.3034	0.2956	0.2940	0.3032	0.3029
		0.4	0	0.3039	0.3007	0.2939	0.2899	0.3023	0.3013
			0.3	0.3045	0.3027	0.2942	0.2928	0.3027	0.3024
10%	0	0	0	0.3075	0.3027	0.2960	0.2901	0.3044	0.3023
			0.3	0.3075	0.3049	0.2960	0.2942	0.3044	0.3038
		0.4	0	0.3070	0.3022	0.2951	0.2893	0.3040	0.3020
			0.3	0.3072	0.3046	0.2950	0.2933	0.3041	0.3036
	0.3	0	0	0.3070	0.3021	0.2961	0.2902	0.3041	0.3021
			0.3	0.3070	0.3043	0.2961	0.2944	0.3041	0.3036
		0.4	0	0.3061	0.3013	0.2949	0.2891	0.3036	0.3016
			0.3	0.3065	0.3039	0.2950	0.2934	0.3038	0.3033
	0.7	0	0	0.3054	0.3005	0.2956	0.2898	0.3033	0.3013
			0.3	0.3053	0.3026	0.2956	0.2939	0.3032	0.3027
		0.4	0	0.3039	0.2990	0.2939	0.2882	0.3023	0.3004
			0.3	0.3045	0.3019	0.2942	0.2926	0.3027	0.3022
20%	0	0	0	0.3075	0.3003	0.2960	0.2879	0.3044	0.2998
			0.3	0.3075	0.3041	0.2960	0.2945	0.3044	0.3027
		0.4	0	0.3070	0.2998	0.2951	0.2871	0.3040	0.2995
			0.3	0.3072	0.3039	0.2950	0.2937	0.3041	0.3025
	0.3	0	0	0.3070	0.2997	0.2961	0.2881	0.3041	0.2995
			0.3	0.3070	0.3035	0.2961	0.2947	0.3041	0.3025
		0.4	0	0.3061	0.2989	0.2949	0.2870	0.3036	0.2990
			0.3	0.3065	0.3032	0.2950	0.2938	0.3038	0.3022
	0.7	0	0	0.3054	0.2982	0.2956	0.2878	0.3033	0.2987
			0.3	0.3053	0.3018	0.2956	0.2943	0.3032	0.3016
		0.4	0	0.3039	0.2967	0.2939	0.2862	0.3023	0.2978
			0.3	0.3045	0.3011	0.2942	0.2931	0.3027	0.3011
30%	0	0	0	0.3075	0.2991	0.2960	0.2861	0.3044	0.2983
			0.3	0.3075	0.3044	0.2960	0.2955	0.3044	0.3028
		0.4	0	0.3070	0.2987	0.2951	0.2855	0.3040	0.2981
			0.3	0.3072	0.3042	0.2950	0.2948	0.3041	0.3026
	0.3	0	0	0.3070	0.2985	0.2961	0.2865	0.3041	0.2981
			0.3	0.3070	0.3039	0.2961	0.2959	0.3041	0.3025
		0.4	0	0.3061	0.2978	0.2949	0.2855	0.3036	0.2976
			0.3	0.3065	0.3035	0.2950	0.2950	0.3038	0.3023
	0.7	0	0	0.3054	0.2970	0.2956	0.2864	0.3033	0.2973
			0.3	0.3053	0.3022	0.2956	0.2956	0.3032	0.3016
		0.4	0	0.3039	0.2956	0.2939	0.2850	0.3023	0.2965
			0.3	0.3045	0.3015	0.2942	0.2945	0.3027	0.3012

Table 9. Monte Carlo Simulation Results: Robustness to Model Misspecification

%Missing	φ_v	φ_u	ρ	IV No Missing $\boldsymbol{w}$ excluded	Three-Step Estimator $\boldsymbol{w}$ excluded	IV No Missing $\boldsymbol{w}^2$ excluded	Three-Step Estimator $\boldsymbol{w}^2$ excluded	IV No Missing $\boldsymbol{w}^3$ excluded	Three-Step Estimator $\boldsymbol{w}^3$ excluded
0%	0	0	0	0.3006	0.3007	0.3006	0.3005	0.3006	0.3005
			0.3	0.3006	0.3007	0.3006	0.3005	0.3006	0.3005
		0.4	0	0.3003	0.3004	0.3003	0.3002	0.3003	0.3002
			0.3	0.3003	0.3005	0.3003	0.3002	0.3003	0.3002
	0.3	0	0	0.3005	0.3007	0.3005	0.3004	0.3005	0.3004
			0.3	0.3005	0.3006	0.3005	0.3004	0.3005	0.3004
		0.4	0	0.3002	0.3004	0.3002	0.3001	0.3002	0.3001
			0.3	0.3002	0.3004	0.3002	0.3002	0.3002	0.3002
	0.7	0	0	0.3003	0.3004	0.3003	0.3002	0.3003	0.3002
			0.3	0.3003	0.3004	0.3003	0.3002	0.3003	0.3002
		0.4	0	0.3000	0.3001	0.3000	0.2999	0.3000	0.2999
			0.3	0.3000	0.3002	0.3000	0.3000	0.3000	0.3000
5%	0	0	0	0.3006	0.3012	0.3006	0.3004	0.3006	0.3006
			0.3	0.3006	0.3020	0.3006	0.3010	0.3006	0.3007
		0.4	0	0.3003	0.3009	0.3003	0.3002	0.3003	0.3003
			0.3	0.3003	0.3018	0.3003	0.3008	0.3003	0.3004
	0.3	0	0	0.3005	0.3011	0.3005	0.3003	0.3005	0.3006
			0.3	0.3005	0.3019	0.3005	0.3009	0.3005	0.3006
		0.4	0	0.3002	0.3009	0.3002	0.3001	0.3002	0.3003
			0.3	0.3002	0.3017	0.3002	0.3007	0.3002	0.3003
	0.7	0	0	0.3003	0.3010	0.3003	0.3002	0.3003	0.3004
			0.3	0.3003	0.3017	0.3003	0.3007	0.3003	0.3003
		0.4	0	0.3000	0.3007	0.3000	0.2999	0.3000	0.3000
			0.3	0.3000	0.3016	0.3000	0.3005	0.3000	0.3001
10%	0	0	0	0.3006	0.3021	0.3006	0.3010	0.3006	0.3005
			0.3	0.3006	0.3034	0.3006	0.3014	0.3006	0.3008
		0.4	0	0.3003	0.3019	0.3003	0.3008	0.3003	0.3002
			0.3	0.3003	0.3031	0.3003	0.3012	0.3003	0.3005
	0.3	0	0	0.3005	0.3021	0.3005	0.3009	0.3005	0.3005
			0.3	0.3005	0.3033	0.3005	0.3014	0.3005	0.3007
		0.4	0	0.3002	0.3018	0.3002	0.3007	0.3002	0.3001
			0.3	0.3002	0.3031	0.3002	0.3012	0.3002	0.3004
	0.7	0	0	0.3003	0.3019	0.3003	0.3007	0.3003	0.3002
			0.3	0.3003	0.3031	0.3003	0.3012	0.3003	0.3005
		0.4	0	0.3000	0.3017	0.3000	0.3005	0.3000	0.2999
			0.3	0.3000	0.3029	0.3000	0.3010	0.3000	0.3002
20%	0	0	0	0.3006	0.3053	0.3006	0.3022	0.3006	0.3002
			0.3	0.3006	0.3078	0.3006	0.3023	0.3006	0.3014
		0.4	0	0.3003	0.3050	0.3003	0.3021	0.3003	0.2999
			0.3	0.3003	0.3075	0.3003	0.3022	0.3003	0.3011
	0.3	0	0	0.3005	0.3052	0.3005	0.3022	0.3005	0.3002
			0.3	0.3005	0.3077	0.3005	0.3023	0.3005	0.3014
		0.4	0	0.3002	0.3049	0.3002	0.3020	0.3002	0.2999
			0.3	0.3002	0.3075	0.3002	0.3022	0.3002	0.3011
	0.7	0	0	0.3003	0.3049	0.3003	0.3020	0.3003	0.3000
			0.3	0.3003	0.3076	0.3003	0.3021	0.3003	0.3011
		0.4	0	0.3000	0.3045	0.3000	0.3019	0.3000	0.2997
			0.3	0.3000	0.3074	0.3000	0.3020	0.3000	0.3009
30%	0	0	0	0.3006	0.3090	0.3006	0.3045	0.3006	0.3004
			0.3	0.3006	0.3127	0.3006	0.3039	0.3006	0.3025
		0.4	0	0.3003	0.3088	0.3003	0.3044	0.3003	0.3001
			0.3	0.3003	0.3126	0.3003	0.3038	0.3003	0.3023
	0.3	0	0	0.3005	0.3091	0.3005	0.3045	0.3005	0.3004
			0.3	0.3005	0.3126	0.3005	0.3038	0.3005	0.3025
		0.4	0	0.3002	0.3089	0.3002	0.3044	0.3002	0.3000
			0.3	0.3002	0.3125	0.3002	0.3038	0.3002	0.3022
	0.7	0	0	0.3003	0.3090	0.3003	0.3045	0.3003	0.3002
			0.3	0.3003	0.3124	0.3003	0.3037	0.3003	0.3023
		0.4	0	0.3000	0.3088	0.3000	0.3043	0.3000	0.2998
			0.3	0.3000	0.3123	0.3000	0.3037	0.3000	0.3020

2.5 Empirical Example

2.5.1 Maternal Prenatal Smoking and Birth Weight

Our empirical application examines the impact of maternal prenatal smoking on birth weight using self-reported data from U.S. birth certificates. Prenatal smoking remains a costly public health concern that is associated with a host of detrimental effects on the health of mothers and infants, including low birth weight, preterm delivery, ectopic pregnancy, spontaneous abortion, placental abruption, sudden infant death syndrome, and congenital disabilities, among others (Dempsey and Benowitz, 2001; Einarson and Riordan, 2009; Lushniak et al., 2014). In addition, complications from prenatal maternal smoking can also adversely affect health outcomes, human capital accumulation, mental development, and well-being later in life (Lushniak et al., 2014; Simon, 2016; Settele and Van Ewijk, 2018). Our empirical application focuses on birth weight as the main determinant of infant health, considering that it is linked to a high risk of infant mortality and substantial cost in terms of hospital charges in the U.S. (Almond et al., 2005).

Estimating the impact of maternal prenatal smoking on birth weight is affected by the presence of plausible unobserved factors that correlate with both prenatal smoking and birth outcomes, leading to biased OLS estimates. To address this endogeneity concern, economic literature on the effect of prenatal smoking relies on cigarette taxes as an instrument for smoking during pregnancy (e.g., see Evans and Ringel 1999, Gruber and Köszegi 2001, Lien and Evans 2005, Brachet 2008,Simon 2016, Settele and Van Ewijk 2018). Following this literature, we exploit federal and state cigarette taxes as instruments for prenatal smoking, assuming that they affect birth weight only through maternal smoking. As reported in Brachet (2008), this assumption is likely to hold in our setting as cigarette tax revenues were not earmarked to fund tobacco control programs during the period of our analysis.

Another potential issue, which is of primary interest in this empirical exercise, is that of missing data on prenatal smoking status in the U.S. Natality data, up to 32.69%, in the immediate years after the 1989 modification to include information about prenatal tobacco use on birth certificates. This combination of endogenous prenatal smoking status that is plausibly MNAR aligns well with our missing data framework and allows us to illustrate the application of our proposed estimator in practice. Our empirical example aims to employ the three-step estimator to account for missing data in estimating the impact of prenatal smoking on birth weight.

2.5.2 Data

Our empirical example employs the publicly available 1989 U.S. Natality data to estimate the impact of maternal smoking on birth weight. The U.S. Natality microdata comes from information on birth certificates and contains data on birth outcomes, health, demographics, and other socioeconomic characteristics of mothers for all births occurring in the U.S. in a given year. Relevant to our empirical application, this dataset contains data on birthweight as well as self-reported information on mothers' tobacco use during pregnancy and the intensity of cigarette use, which started being included in the birth certificate in 1989. We rely on the mother's tobacco use during pregnancy to generate an indicator of prenatal smoking status, which takes the value of one if a mother reported using tobacco during pregnancy and zero otherwise, and focus on 1989 data as it contains the highest prevalence of missing data on smoking status, approximately 32.69%, compared to other years. We rely on two variables as potential drivers of missingness: first, attendant present at birth, an indicator variable which takes the value of one for Doctor of Medicine (M.D.) or Doctor of Osteopathy (O.O), and zero for midwives and others, and second, an indicator for birth certificate reporting requirements of the state or Standard Metropolitan Statistical Area (SMSA) of residence, which takes the value of one if at least one item was not reported in the birth certificate and zero if the state reported all items in the birth certificates. We also include other health, demographic, and socioeconomic control variables such as sex of the child, gestation, Kessner index, live-birth order, and mother's age, marital status, educational attainment, and race. In addition, the publicly available U.S. Natality data before 2005 contains geographic variables, including mother's states of residency and the state where they gave birth. This information allows us to merge this data with other state-level policies, including cigarette taxes.

Table 10. Summary Statistics of Birth Weight, Tobacco Use, Cigarette Taxes, and Demographic Characteristics, 1989 U.S. Natality Data from the National Vital Statistics System.

Statistic	Mean	St.Dev.
Birth weight	3341.649	602.847
1(Tobacco use in pregnancy)	0.194	0.396
Cigarette tax	0.391	0.096
Mother's age	26.284	5.681
Married	0.729	0.445
Mother's race:		
White	0.786	0.410
Black	0.171	0.376
Other	0.043	0.203
Mother's education:		
0 - 8 years	0.058	0.234
9 - 11 years	0.172	0.377
12 years	0.391	0.488
13 - 15 years	0.204	0.403
16 years and over	0.175	0.380
Girl birth	0.488	0.500
Birth order:		
First child	0.411	0.492
Second child	0.324	0.468
Third child and over	0.265	0.442
Kessner index:		
Adequate	0.658	0.474
Intermediate	0.227	0.419
Inadequate	0.086	0.281
Unknown	0.028	0.165
Gestation		
Under 37 weeks	0.106	0.308
37 - 39 weeks	0.407	0.491
40 weeks	0.224	0.417
41 weeks	0.143	0.350
42 weeks and over	0.120	0.325

We obtained state and federal cigarette tax data from the Centers for Disease Control and Prevention (CDC) data on Tax Burden on Tobacco. Following the literature that links cigarette taxes and prenatal smoking, we merged the state and federal cigarette taxes with U.S. Natality based on the state where the birth occurred (Evans and Ringel, 1999; Brachet, 2008). Note that there is little variation between residence and birth states in the data. For instance, in the U.S. 1989 Natality data, birth and residence states are the same in about 97.44% of births. Table 10 provides descriptive statistics of variables used in our empirical application, including birth weight, tobacco use, cigarette taxes, and other health and Demographic Characteristics in the 1989 U.S. Natality data. Approximately 19.4% of mothers smoked during pregnancy in our sample, and the average birth weight is around 3341.649 grams.

2.5.3 Results

We present regression results from our empirical example in Table 11. The first two columns, "OLS-CC" and "IV-CC," present results from complete case analysis OLS and IV estimators, and the last column "Three-Step Estimator," presents estimates from our proposed method, which account for missing data in prenatal smoking status. Our complete-case analysis OLS estimates suggest that prenatal smoking is associated with a statistically significant 205.71 grams reduction in birth weight. These OLS results are plausibly biased due to unobservable variables that may be correlated with prenatal smoking and birth weight. Using state and federal cigarette tax as an instrument for prenatal maternal smoking, our IV estimates in the second column show that smoking significantly lowers birth weight by 431.19 grams, approximately twice the magnitude of our OLS estimate. In the presence of ideal data, for instance, if prenatal smoking is MCAR and free from misreporting, these results suggest that OLS estimates are upward biased and a counterintuitive relationship between birthweight and unobserved factors that are associated with smoking. Thus, unobservable factors positively associated with smoking improve birth weight, or unobserved behaviors that lower prenatal smoking are associated with lower birth weight.

Table 11. Effect of Prenatal Maternal Smoking on Birth Weight, 1989 U.S. Natality Data from the National Vital Statistics System.

	OLS-CC	IV-CC	Three-Step Estimator
1(Tobacco use in pregnancy)	-205.706 ***	-431.193 ***	-172.773***
	(0.837)	(34.420)	(5.611)
Mother's age	-0.306***	0.418***	-2.599 ***
	(0.074)	(0.133)	(0.126)
Married	33.361***	4.953	18.079 ***
	(0.894)	(4.429)	(1.536)
Mother's race:			
White	94.478***	119.998***	106.521***
	(1.837)	(4.317)	(1.499)
Black	-77.739***	-83.389***	-26.453 ***
	(1.979)	(2.184)	(2.201)
Mother's education:			
0 - 8 years	-71.995***	-45.058***	-59.707 ***
	(1.834)	(4.512)	(1.453)
9 - 11 years	-67.787***	-6.591	-51.326 ***
	(1.315)	(9.433)	. (2.044)
12 years	-29.243***	7.361	-23.585***
	(0.995)	(5.676)	(1.379)
13 - 15 years	-8.597***	10.524***	-7.125 ***
	(1.053)	(3.107)	(1.093)
Girl birth	-134.333***	-134.374***	-132.669 ***
	(0.635)	(0.643)	(0.448)
Birth order:			
First child	-111.712***	-124.587***	-100.434 ***
	(0.902)	(2.167)	(0.824)
Second child	-27.848***	-31.816***	-25.249 ***
	(0.867)	(1.067)	(0.748)
Kessner index:			
Adequate	94.054***	89.909***	96.358 ***
	(2.082)	(2.203)	(1.582)
Intermediate	66.224***	67.365***	71.805 ***
	(2.145)	(2.182)	(1.528)
Inadequate	12.584***	18.838***	11.831 ***
	(2.335)	(2.552)	(1.817)
Gestation:			
Under 37 weeks	-918.879***	-916.456***	-900.775 ***
	(1.349)	(1.417)	(1.296)
37 - 39 weeks	-238.336***	-242.018***	-232.221 ***
	(1.047)	(1.201)	(0.824)
Gestation - 40 weeks	-46.034***	-51.699***	-42.972 ***
	(1.137)	(1.441)	(0.857)
41 weeks	30.092***	25.352***	31.776 ***
	(1.243)	(1.453)	(0.979)

Note: Standard errors in parenthesis and bootstrapped for the three-step estimator. The analytical sample comes from the 1989 U.S. Natality Data and constists of all births reported in the U.S. Coefficients of the constant term not reported. There are 2,607,019 observations in OLS-CC and IV-CC columns and 3,625,568 observations in the three-step estimator. Significance codes: *** p<0.01, ** p<0.05,* p<0.1

However, prenatal smoking status is plausibly not MCAR, as literature documents a potential link between smoking status and the probability of reporting it in data (Jackson et al., 2014). Note that U.S. Natality data is based on self-reported smoking status, which may bring concerns about measurement error in prenatal smoking. As Brachet (2008) estimates suggest, this misreporting is mostly one-sided (false negative rather than false positive), and our framework partially accounts for this type of misreporting by converting the missing data problem to that of underreporting. Using our three-step estimator in column three, we find prenatal smoking leads to a statistically significant 172.77 grams reduction in birth weight. These estimates are 19% lower in magnitude than OLS estimates, providing suggestive evidence for a more likely argument that unobserved factors that are associated with prenatal smoking negatively affect birth weight or those that are negatively correlated with smoking increase birth weight. In addition, these estimates are almost 150% lower in magnitude than the conventional IV estimates.

2.6 Conclusion

Our paper examines the estimation of treatment effects when participation is missing and plausibly endogenous. We derive the expressions for asymptotic bias in complete-case analysis OLS and IV estimators and discuss the bias under different missing data mechanisms, including the general case of MNAR. In particular, complete-case OLS and IV estimators can be affected by either upward or downward bias when treatment is not MCAR, and in some settings, the estimated treatment effects can have an opposite sign to the true effects. Not accounting for missing data in such settings can result in erroneous conclusions about the impact of programs, which can plausibly mislead policy decisions.

We propose a three-step estimator and show that it consistently recovers treatment effects of interest. Our proposed estimator relies on the availability of an instrument for participation that is not correlated with missingness and an additional variable associated with missing data. In Monte Carlo Simulations, we demonstrate that the proposed approach performs well in finite samples and is superior to complete-case analysis when data is not MCAR. Finally, we illustrate the application of our three-step estimator in practice by employing it to estimate the effect of maternal prenatal smoking on birth weight using U.S. Natality data. Note that our current framework does not address other plausible data issues in evaluating the impact of programs, such as bidirectional misreporting in binary treatment variables, which can coexist with the problem of missing data. This limitation of our study is a potential area that future research can explore, as misreporting in binary regressor is prevalent in survey data. Our paper also imposes distributional and functional form assumptions in estimating treatment effects of interest. Future work should consider approaches that relax these assumptions.

Chapter III: Air Pollution and Student Achievement: Evidence from Tanzania[1]

3.1 Introduction

Many developed and developing countries use student performance on high-stakes exams for teacher evaluation, fund allocations, and for measuring relative performance and developments over time. Students' performance on these exams also determines their course of future education and, consequently, their labor market trajectory. Given these, it is important that inequalities in testing conditions across space and over time are factored in the aforementioned policies and labor market allocations to reduce inequalities and inefficiencies. In this paper, we explore students' performance on high-stake exams in Tanzania and examine whether the variation in air pollution exposure on the day exam across space is a source of unfairness in testing conditions, noise on students' abilities, and inefficiencies in education policies and labor market allocations.

Our paper leverages spatial and temporal variation in air pollution exposure on examination day to examine the impact of contemporaneous air pollution exposure on students' performance in high-stakes exams in Tanzania. We use national secondary school leaving examinations results in Tanzania, administered by the National Examination Council of Tanzania (NECTA) across all public and most private schools. We combine the rich nationwide administrative datasets on secondary school leaving examination results, containing a universe of students sitting for these exams, with data on examination center characteristics, atmospheric conditions, and satellite reanalysis data on exposure to fine particulate matter, PM2.5. To causally identify the effect of air pollution on exam performance, we employ plausibly exogenous changes in local wind direction to instrument for air pollution exposure of students in an instrumental variables approach (Deryugina et al., 2019).

Our findings indicate a large negative effect of exposure to air pollution on the day of the exam on a student's exam performance. Specifically, in our preferred specification, student-level IV estimations, an increase in $PM2.5$ concentration by $10\ \mu g/m^3$ on the day a student appears for the exam worsens their performance on the exam by 0.05 standard deviations. This result is

[1]This paper is a joint work with Tejendra Pratap Singh (Department of Economics, Georgia State University, Email: tpratapsingh1@gsu.edu).

equivalent to a worsening of teacher value-added by approximately 0.33 standard deviations (Bau and Das, 2020). We establish the robustness of our results through a series of falsification checks. Additionally, we carry out a heterogeneity analysis and find that the effects are more pronounced for female students, younger students, students who appear for exams in government-owned examination centers, students in relatively poorer regions, and those at the lower end of the achievement distribution. Leveraging the information on students' exam scores on multiple subjects, we find evidence that our results could be driven by worsening performance in the subjects that test fluid intelligence. Our findings further indicate that the effects on student performance are more pronounced on extremely polluted days. We document no extensive margin effect of whether the student appears or not for the exam on more polluted days. However, the likelihood of students failing an exam goes up if they appear for an exam on a more polluted day.

Previous literature shows that exposure to air pollution has a detrimental effect on cognitive performance, often mediated through adverse health effects emanating from the exposure to worse air quality (Dockery and Pope, 1996; Pope III and Dockery, 2006; Schikowski and Altuğ, 2020). Studies in developed countries find that exposure to air pollution lowers cognitive performance among the elderly (Ailshire and Clarke, 2015; Wilker et al., 2015; Bishop et al., 2018), among adults in prime working age (La Nauze and Severnini, 2021), and lowers students' exams performance (Ebenstein et al., 2016; Cho, Forthcoming). Studies in developing countries similarly find short- and long-term exposure to air pollution impedes students' cognitive performance (Graff Zivin et al., 2020; Balakrishnan and Tsaneva, 2021; Bedi et al., 2021; Carneiro et al., 2021).

We make multiple contributions to the earlier literature examining the impact of air pollution exposure on student performance. First, we build on previous work, which has almost exclusively focused on a small geographical area, and provide the first nationwide estimates of the impact of air pollution exposure on student performance. Second, we contribute to the existing work by providing estimates of this relationship for Africa, a region that has not received attention from existing work until now. Evidence from the previous studies may not necessarily extend to this context considering the documented non-linear relationship between air pollution exposure and education outcomes and potential differences in the intensity of air pollution exposure, avoidance costs, and the stringency of environmental regulations. Third, we also provide novel evidence on the heterogenous effects of air pollution exposure on student performance by leveraging our large sample size. Finally, we also contribute to the large literature examining the impact of air pollution exposure on various aspects of human capital accumulation (Aguilar-Gomez et al., 2022).

The rest of the paper is organized as follows. Section 3.2 gives a background of our setting and discusses the related literature. Section 3.3 provides a description of data and presents descriptive statistics. Section 3.4 explains empirical strategies that we employ. Section 3.5 presents results. Section 3.6 contextualizes our main results and concludes.

3.2 Background

While reviewing the education sector in Tanzania is beyond the scope of this paper, we provide a brief overview of secondary school education in Tanzania. Interested readers can refer to Yusuph (2019) for a detailed description of the education sector in Tanzania. Our paper also builds on the existing literature on air pollution, cognitive performance, and human capital accumulation. In this section, we discuss the existing body of works on these topics and highlight our contribution.

3.2.1 Secondary Education in Tanzania

The education system in Tanzania consists of six years of secondary education, with four years at the Ordinary Level (O-Level) and two years at the Advanced Level (A-Level). The majority of secondary schools in Tanzania are government-owned. For instance, in 2017 and 2018, about 74.4.% and 74.8% of secondary schools were government-owned (MoEST, 2019). Secondary education has been free in all government-owned schools since the implementation of the fee-free secondary school policy in January 2016.

Admission to secondary school is based on the nationwide Primary School Leaving Examination (PSLE), which is carried out after seven years of primary education. In the second year of secondary schooling, students take a national exam, Form Two National Assessment (FTNA). Students who clear this exam can continue for two more years of secondary education, and those who did not pass the exam can re-sit for the following year. At the end of O-Level, students appear in a nationwide assessment, the Certificate of Secondary Education Examination (CSEE). All students in secondary schools with four years of education and FTNA passing grades are eligible to enroll for CSEE. Students who passed FTNA or its equivalent and those who previously sat for CSEE are also eligible to appear as private candidates in registered examination centers.

A candidate must register for the core and compulsory subjects (civics, Kiswahili, English language, biology, and mathematics). In addition, each candidate can register for not less than two additional subjects (natural science, social science, technical subjects, agriculture, commercial, home economics, arts, computer Studies, or foreign languages) and not more than one optional subject (religious subjects). Each subject has a separate examination either on different days or in two distinct sessions of a day. Each subject exam is between two and a quarter to three and a half hours long and contains a mix of objective and subjective questions.

ACSEE exams are taken by all students who have completed two years of A-Level secondary education and have secured three credits (A, B+, B, or C) at the CSEE and the subject combination. Students are required to register for the compulsory General Studies examination in addition to a maximum of five other subjects depending on the combination of subjects studied at the A-Level. Each subject examination duration is at least three hours. For some subjects, the theory and practical parts of the examination are spread over distinct days. Like CSEE, ACSEE subject examinations are administered in two separate sessions on each examination day. Students registered for the same subject sit for the exam in the same session on the same day.

All national examinations, including CSEE and ACSEE, are administered centrally by NECTA. This government institution generally oversees the registration of candidates and examination centers; sets, moderates, and distributes national exams to the centers; provides external supervisors and invigilators through regional committees; supervises their marking; and publishes results. The Examinations Regulations (2015) provides provisions for nationwide assessments of students, including those discussed earlier. Students are expected to appear for national examinations where they are registered and not allowed to transfer to a different examination center without the approval of the NECTA executive secretary once the registration period is over. Additionally, exams are expected to continue if they fall on public holidays unless otherwise stated by NECTA, and, on average, there is a gap of two to five months between the registration end date and the public release of the examination schedule. Early registration, restrictions on changing exam center, and fixed exam dates makes self-selection on exam dates and location implausible.

The stakes of missing or failing CSEE and ACSEE are considerably high. It is only in the presence of sufficient proof that a student who fails to appear for the exam, and upon approval from the ministry responsible for education, may be permitted to sit for the exam the following year. If approved, their examination results which may be needed for the next academic level are issued in the year of re-sitting, creating a gap in their education progression. In addition, national examinations at the end of each education cycle determine the progression to higher levels of education. For instance, CSEE exam scores and completion of two years of A-Level education are the main criteria determining qualification to register and sit for the ACSEE. Performance ACSEE, in combination with CSEE in most cases, is used to admit students in higher education (like colleges and universities) and specific courses. Hence, CSEE and ACSEE assessments can potentially determine students' education attainment, career trajectory, and lifetime earnings.

3.2.2 Related Literature

Exposure to particulate matter is associated with adverse health effects, including short and long-run cognitive impairment. Fine inhalable particulate matter, $PM2.5$, can penetrate the respiratory systems deep in the lungs and be absorbed in the bloodstream or trans-locate to other organs, for instance, the brain, indirectly through the bloodstream or directly through the nose (Oberdörster et al., 2004; Pope III and Dockery, 2006; Peters et al., 2006; Kampa and Castanas, 2008; Mills et al., 2009). Consequentially, this may lead to lower quality of oxygen in the blood circulation, inflammatory responses in the brain, and central nervous system disorders such as stroke, cerebrovascular disease, and cognitive impairment (Pope III and Dockery, 2006; Kampa and Castanas, 2008; Calderón-Garcidueñas et al., 2008; Schikowski and Altuğ, 2020), which may in turn affect cognitive function.

Studies that explore the link between air pollution exposure and cognitive functions provide evidence that exposure to air pollution lowers cognitive functions across different population groups. Accounting for individual and neighborhood characteristics, Ailshire and Clarke (2015) find that older adults living in areas with high $PM2.5$ concentrations had a significantly higher error rate than those exposed to lower concentrations. Similarly, Wilker et al. (2015) find exposure to $PM2.5$ impacts brain aging amongst older adults, even for those who are dementia and stroke-free, and Bishop et al. (2018) find exposure to $PM2.5$ increases the risk of dementia amongst elderly. Using data from brain-training games in the U.S., La Nauze and Severnini (2021) instrument $PM2.5$ exposure with daily changes in wind direction and find that exposure to air pollutants reduces adult cognitive performance with a larger effect on prime working-age adults, those with a lower ability, and on the new tasks. Bedi et al. (2021) provide experimental evidence on the effect of short-term exposure to $PM2.5$ on a range of cognitive domains at a university in Brazil using 54 lab sessions over three years and find that high levels of $PM2.5$ reduce performance on fluid reasoning tests.

Our interest in this paper is in estimating the causal effects of exposure to air pollution, $PM2.5$, on performance in high-stakes exams. In this way, our study is mostly related to the literature on the impact of air pollution on education outcomes. In Israel, Ebenstein et al. (2016) examine the impact of transitory, plausible exogenous exposure to $PM2.5$ on students' performance in high-stakes compulsory exams and post-secondary outcomes. The study employs fixed-effects models and finds that transitory $PM2.5$ exposure during exams reduces students' academic performance, overall education attainment, and earnings. Using data on nearly 20 counties in the U.S., Marcotte (2017) shows that fine airborne particulate matter exposure on days of exams lowers children's math and reading scores. Heissel et al. (2022) exploit variations in wind direction and student school switching near major highways in Florida to estimate the effect of traffic pollution on student outcomes. The study finds that absence rates and behavioral incidents increase while test scores decrease when the students attend a downwind school relative to an upwind school. Persico and Venator (2021) use proximity to Toxics Release Inventory sites in Florida and provide evidence that air pollution is associated with lower test scores and a higher probability of suspension from school. In Chicago, Komisarow and Pakhtigian (2022) find coal plant closure improves school attendance.

Our study also adds to the recent growing literature that examines the link between air pollution exposure and learning in developing countries. In Santiago, Chile, Bharadwaj et al. (2017) find that utero exposure to air pollution leads to lower math and language skills test scores in the fourth grade. Graff Zivin et al. (2020) use random variation in wind direction to identify the causal effect of agricultural fires on cognitive performance in a high-stakes exam setting in China. Their findings suggest that an increase in the difference between the upwind and downwind fires during the National College Entrance Examination lowers the exam score and the probability of getting into the first-tier universities. Using plausibly exogenous spatial and temporal variation in air pollution and an instrumental variable approach using wind direction as an instrument, Carneiro et al. (2021) find that contemporaneous $PM10$ exposure on national-wide university entrance exams in two most industrious states in Brazil, Rio de Janeiro and Sao Paulo, lowers student performance. Balakrishnan and Tsaneva (2021) use nighttime thermal inversion as an instrumental variable for exposure to particulate matter and estimate the effect of air pollution on reading and math scores in rural India.

Evidence from these studies may not necessarily extend to other countries, considering that air pollution in developed countries is relatively low and the documented plausible non-linear relationship between air pollution and exam outcomes. Furthermore, economic activities, air quality regulations, implementation and stringency, population responses to air pollution, and air pollution avoidance costs vary considerably between developed and developing countries (Arceo et al., 2016). Though evidence in the literature suggests that air pollution lowers education performance, there is little evidence from Sub-Saharan Africa (Amegah and Agyei-Mensah, 2017; Coker and Kizito, 2018). We contribute to the existing literature by estimating the relationship between air pollution exposure and exam performance in Tanzania. In addition, our study provides the first nationwide estimates of the impact of air pollution exposure on students' performance, as earlier literature has primarily examined small geographical areas. Our large sample size of the universe of students sitting for national exams in Tanzania also allows us to explore and discuss the heterogeneous effects of air pollution exposure on exam performance. Finally, our study contributes to the broad literature on the impact of air pollution exposure on human capital accumulation.

3.3 Data and Summary Statistics

The ideal data to uncover the impact of air pollution on student achievement would contain a measure of each student's air pollution exposure, their exam scores, and other potential confounders that affect both air pollution exposure and exam scores. While we do not have access to such data, we use multiple data sources to come close to these ideal data. We detail each data source below.

3.3.1 Student Test Scores Data

We use data on student achievement made publicly available by The National Examinations Council Of Tanzania (NECTA). NECTA is a government agency that administers national tests in Tanzania. NECTA provides test scores for the universe of students who enroll for various national tests. We use the information on student test scores for two national exams – CSEE and ACSEE. For both CSEE and ACSEE examinations, we observe a student's aggregate score, score in each subject they enrolled in, and sex of the student. Unfortunately, apart from the sex of the student, we do not have information on any other demographic. We also do not observe student performance on different exam components for any subject, both for CSEE and ACSEE. We scrape student test scores data for CSEE and ACSEE from 2016-2019.

Scoring on CSEE is subject-specific, and these subject-specific scores are aggregated to determine whether a student passes the exam or not. Subject-specific scoring is based on a five-point scale, with the higher points corresponding to worse performance on the exam. Since students need to pass seven subjects in order to be awarded one of the four divisions, the aggregate score can range from seven to thirty-five. These divisions are awarded in decreasing order of overall exam performance, i.e., the first division represents performance at the top end of the exam performance distribution. In contrast, the fourth division represents performance at the bottom end of the exam performance distribution.[2]

Scoring on ACSEE is also subject-specific. Like CSEE, ACSEE subject-specific scores are aggregated to determine whether students clear ACSEE exams or not. Scoring on ACSEE is on a seven-point scale, with increasing points representing worse performance on the exam. Since students are required to pass three principle subjects, the aggregate score for ACSEE exams can range from three to twenty-one. The compulsory General Studies exam is administered at the ACSEE level to determine whether the student scores in the two top divisions.

3.3.2 Air Pollution Data

We use air pollution data from CAMS-EAC4 (ECMWF Atmospheric Composition Reanalysis 4) (Inness et al., 2019). These data come from a global reanalysis of atmospheric conditions. Reanalysis combines model data with observations from across the world into a globally complete and consistent dataset using a model of the atmosphere based on the laws of physics and chemistry. These data are available at a spatial resolution of $0.1° \times 0.1°$ and a temporal resolution of three hours.

We use data on $PM2.5$ concentration to construct a measure of air pollution exposure of students. $PM2.5$ are suspended particles that are less than $2.5\ \mu m$ in diameter. Due to their small size the human body is often unable to filter them out effectively. Therefore, they are able to travel deep into organs such as the heart and enter the bloodstream. It should be noted that the satellite measure of $PM2.5$ concentration is an estimate of the true ground-level pollution concentration. Previous work has shown that these data correlate well with the pollution concentration measured by the Environmental Protection Agency (EPA) ground monitors (van

[2] Even though the student did not pass the CSEE exam, they can proceed to the fifth year of secondary school if they score more than forty-five out of a hundred on at least five subjects in CSEE.

Donkelaar et al., 2021). However, some work has shown that these satellite reanalysis data might underestimate $PM2.5$ concentration for high concentration levels (Fowlie et al., 2019). To the extent that students' performance on exams on more polluted days worsens, our results can be interpreted as a lower bound on the true effect.

3.3.3 Weather Data

We use weather data from the ERA5-Land. ERA5-Land is a reanalysis dataset providing a consistent view of the evolution of land variables over several decades. Reanalysis produces data that goes several decades back in time, providing an accurate description of the past climate. ERA-5 satellite data on climatic conditions is found to be highly correlated and is considered a good proxy for surface climatic conditions (Gleixner et al., 2020; Muñoz Sabater et al., 2021; Ssenyunzi et al., 2020). These data are available at the spatial resolution $0.1° \times 0.1°$ ($10km \times 10km$ at the equator). Temporally these data are available every hour. We use data on weather conditions that might be correlated with both air pollution and student achievement. Specifically, we use data on cumulative precipitation, temperature, wind speed, dew point (a measure of humidity), and atmospheric pressure.

3.3.4 Analytical Sample Construction

In order to geocode examination centers, we rely on two sources. First, we use the information on the geolocation of secondary schools provided by the government of Tanzania. Second, to geocode the remaining examination centers that are not reflected in the government data, we use Google Maps API. This allows us to geocode all examination centers in our data. We use gridded data on air pollution and weather to assign each examination center to the grid with which it intersects. For air pollution, this allows us to obtain information on $PM2.5$ concentration every three hours, and for weather conditions, we are able to obtain information every hour for each examination center. As we discussed in section 3.3.1, some of the examinations are spread over multiple days. We take an average of $PM2.5$ concentration and weather conditions on all days for which the exam for a particular subject is scheduled. This allows us to create a panel of air pollution and weather conditions for each examination center and subject in which students can enroll. Finally, we match this panel to data on student achievement using subject and examination center information for each student.

3.3.5 Summary Statistics

We present summary statistics from our analytical sample in Table 12 and Table 13. Table 12 presents summary statistics from the exam-level analytical sample. A few things stand out from these summary statistics. First, for CSEE exams, the average student has an aggregate score that would put them in the lowest score division. Second, the performance in mathematics subjects is on average worse than in the reading subjects. Third, in all these scores, there is substantial variation across students. Fourth, average $PM2.5$ exposure is larger than what is observed in developed countries (Gilraine and Zheng, 2022). Table 13 presents summary statistics from the student-level analytical sample. We see that the enrollment of students across gender is roughly equal. We also see that majority of students are enrolled in government-owned schools.

We also show the locations of examination centers in Figure 2. We conclude that examination centers are geographically spread out in Tanzania, as one would expect from the universe of examination centers. We also plot the distribution of $PM2.5$ concentration during our sample period over the spatial region that we study in Figure 3. As can be seen in the figure, there is large spatial heterogeneity in air pollution exposure over the region that we study. In our empirical strategy, we leverage this spatial variation in air pollution exposure to look at the effect of air pollution exposure on student achievement.

3.4 Empirical Strategy

In order to uncover the causal effect of air pollution exposure on student achievement, we employ fixed-effects and IV strategies. We discuss both strategies in detail below.

3.4.1 Fixed-Effects Estimation

We start with estimating the following fixed-effects specification.

$$y_{is} = \alpha_i + \rho_s + \mathrm{W}_{e(is(t))}\boldsymbol{\gamma} + \beta PM2.5_{e(is(t))} + \epsilon_{is} \tag{22}$$

Table 12. Summary Statistics - Exam-Level Data

	N	Mean	SD
(A) CSEE Test Scores			
Aggregate Score (7-35)	12,147,061	28.034	5.585
Mathematics Score (1-5)	2,043,141	4.635	0.716
Reading Score (1-5)	6,647,955	3.981	0.927
(B) ACSEE Test Scores			
Aggregate Score (3-21)	1,346,473	12.753	2.921
Mathematics Score (1-7)	287,247	5.191	1.486
Reading Score (1-7)	359,596	3.930	1.124
(C) Pollution			
$PM2.5\ (\mu g/m^3)$	13,916,070	11.507	10.419
(D) Weather			
Dewpoint Temperature (K)	13,916,070	289.678	3.917
Temperature (K)	13,916,070	296.238	2.743
Surface Pressure (mmHg)	13,916,070	90215.584	6001.885
Total Precipitation (m)	13,916,070	0.002	0.003
Wind Speed (m/s)	13,916,070	2.047	1.170

Notes: Each observation corresponds to a unique subject and is at the student level. Panel A and Panel B presents summary statistics of data from the The National Examinations Council of Tanzania (NECTA). Panel C presents summary statistics of data from the EAC4 (ECMWF Atmospheric Composition Reanalysis 4). $PM2.5$ concentration is aggregated to 24 hour measure. Panel D presents summary statistics of data from the ERA5-Land. Weather variables are aggregated to 24 hour measure. The sample contains data on exams conducted from 2016 to 2019.

where y_{is} is the score of student i in the subject s that they enrolled for, α_i are student fixed-effects that control for any time-invariant student-level unobserved characteristics, ρ_s are subject fixed-effects that control for subject specific time invariant unobserved characteristics, $\mathrm{W}_{e(is(t))}$ is a vector of atmospheric conditions at examination center e where student i appear for the exam of subject s on date $s\ (t)$ that could affect both pollution and student achievement, $PM2.5_{e(is(t))}$ is the $PM2.5$ concentration on the day of the exam at examination center e where student i took the exam, and ϵ_{is} is an error term that we cluster at the examination center level. For atmospheric conditions, we control for cumulative precipitation, temperature, dew point

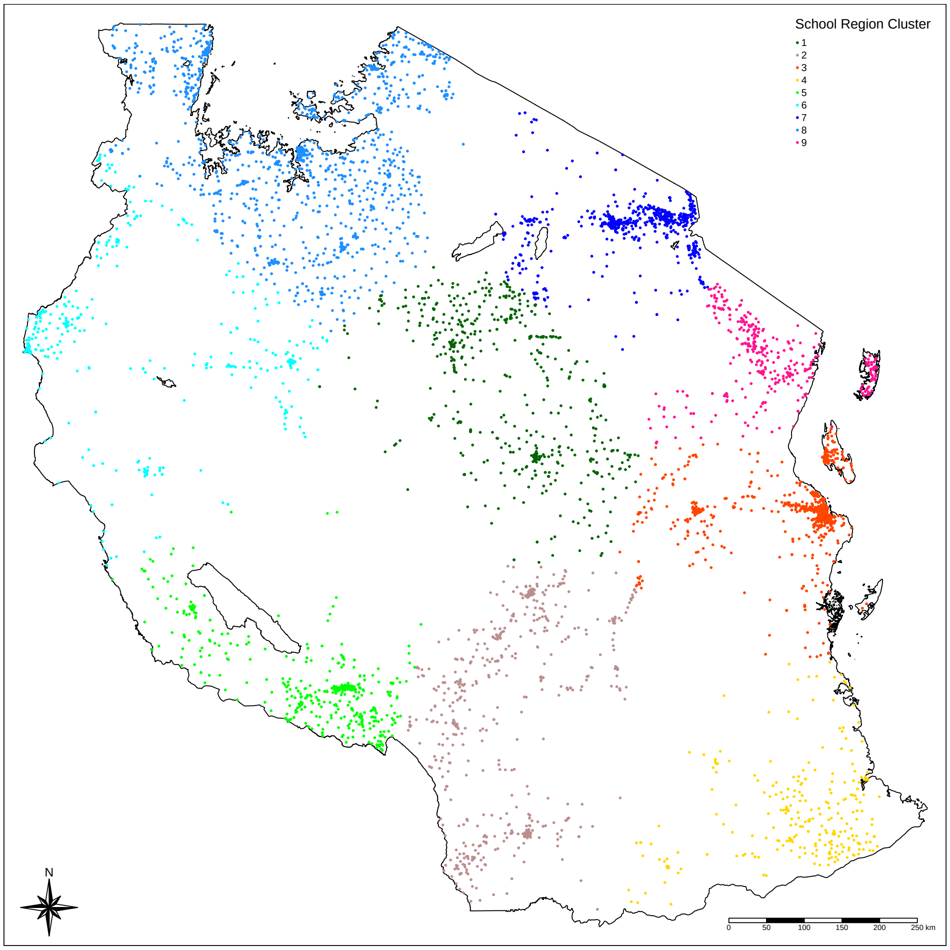
Table 13. Summary Statistics - Student-Level Data

	N	Mean	SD
(A) Division (CSEE)			
Division I	1,483,749	0.032	0.176
Division II	1,483,749	0.100	0.300
Division III	1,483,749	0.160	0.366
Division IV	1,483,749	0.480	0.500
Fail	1,483,749	0.228	0.420
(B) Division (ACSEE)			
Division I	304,244	0.113	0.316
Division II	304,244	0.410	0.492
Division III	304,244	0.406	0.491
Division IV	304,244	0.048	0.213
Fail	304,244	0.023	0.151
(C) Sex			
Male	1,787,993	0.505	0.500
Female	1,787,993	0.495	0.500
(D) Examination Centre Ownership			
Examination Centre Government Owned	1,543,854	0.792	0.406
Examination Centre Private Owned	1,543,854	0.208	0.406
(E) Exam Level			
ACSEE Exam	1,787,993	0.170	0.376
CSEE Exam	1,787,993	0.830	0.376

Notes: Each observation corresponds to a unique student. All panels presents summary statistics of data from the The National Examinations Council of Tanzania (NECTA). The sample contains data on exams conducted from 2016 to 2019.

temperature, sea level pressure, and wind speed. Following Deryugina et al. (2019), we control for these atmospheric conditions non-parametrically. For each variable, we generate indicators for the variable falling in one of four quartiles of the sample distribution. This allows us to control for the potential non-linear relationship between weather and student achievement. Our coefficient of interest is β which represents the change in student score for a $1\mu g/m^3$ increase in $PM2.5$ concentration, on average. For this marginal effect to have a causal interpretation, we

Figure 2. Examination Center Locations

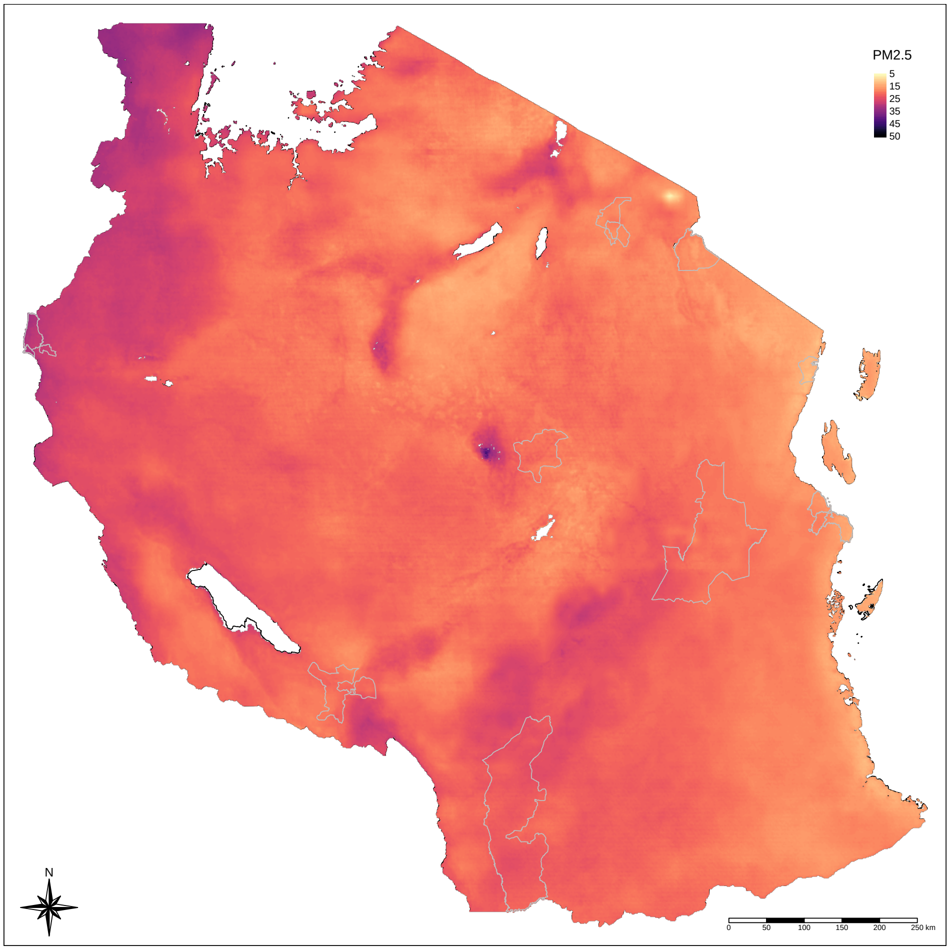


Note: This figure plots the location of all examination centers with geolocation information. Each dot represents a unique examination center. Each examination center is colored based on the cluster it is assigned to using k-means clustering algorithm.

need $PM2.5$ concentration to be uncorrelated with all the unobserved characteristics affecting student exam scores.

While our student fixed-effects control for all such time-invariant unobserved characteristics, we might fail to control for unobserved characteristics that are time-varying. We discuss ways we overcome this challenge below. An additional concern with this specification is the use of student fixed-effects. It is plausible that after controlling for these, our identifying variation is severely restricted. To address this concern, we present the distribution of standard

Figure 3. $PM2.5$ Distribution



Note: This figure plots the distribution of $PM2.5$ concentration in Tanzania during our sample period.

deviation of $PM2.5$ concentration for students in our sample. We present this in Figure E1. We see that there is considerable variation in within student $PM2.5$ concentration measure.

3.4.2 Instrumental Variables Estimation

In addition to the omitted variables issue discussed in section 3.4.1, there is an additional concern regarding the measurement error in air pollution exposure of students on the day of the exam. Since we do not have information on students' residence, there may be a classical measurement error in our $PM2.5$ concentration measure. In such a scenario, our fixed-effects estimates would be attenuated.

In order to overcome the problems of measurement error and omitted variable bias, we instrument $PM2.5$ concentration by leveraging the plausibly exogenous variation in local wind direction (Anderson, 2019; Deryugina et al., 2019; Heissel et al., 2022). Following Deryugina et al. (2019), we first determine the optimal number of clusters in which examination centers are grouped based on their locations using k-means clustering. This exercise, after efficiently trading off bias and variance, yields nine clusters.[3] Our first stage specification is as follows.

$$PM2.5_{e(is(t))} = \alpha_i + \rho_s + \mathrm{W}_{e(is(t))}\boldsymbol{\mu} + \sum_{g=1}^{G}\sum_{b=2}^{B} \nu_{\{g,b\}} \mathbb{1}\left(K_e = g\right) \times \mathbb{1}\left(WinDir_{e(is(t))} = b\right) + \xi_{is} \tag{23}$$

where $\mathbb{1}\left(K_e = g\right)$ is an indicator variable that turns on if examination center e is assigned to cluster g and takes a value zero otherwise, $\mathbb{1}\left(WinDir_{e(is(t))} = b\right)$ is an indicator variable that turns on if the direction from where wind is blowing is in wind direction bin b at examination center e where student i appear for the exam of subject s on date $s\,(t)$, and ξ_{is} is an idiosyncratic error term which we cluster at examination center. Every other parameter in equation (23) is same as in equation (22). We use seven octants to classify the direction from which the wind is blowing. These are $45°$ intervals from $45°$ to $360°$. We omit interval $[0°, 45°\,)$ and use it as the reference wind direction. $\nu_{\{g,b\}}$, which reflects how air pollution changes when the wind is blowing from a particular direction, is allowed to differ across different clusters.

By limiting the effect of a particular wind direction to be the same across all the examination centers in a cluster, we reduce the influence of local pollution sources, which reduces measurement error in our specifications. Since non-local sources of air pollution are likely to exert a similar influence on pollution at all examination centers in a cluster, they drive pollution variation in our first stage (Deryugina et al., 2019). Our IV strategy allows us to abstract away from knowing the precise source of pollution by leveraging changes in local wind direction that

[3] Since k-means clustering is sensitive to the maximum number of clusters chosen by the researcher, we restricted the maximum number of groups to be 31, which is equivalent to the number of administrative regions in Tanzania. In Figure 2, we color each examination center depending on its resulting cluster.

are arguably uncorrelated with unobserved characteristics that affect air pollution exposure and student achievement. In order to validate our IV strategy, we report the first stage F-statistic in all specifications where we employ the IV estimator. As for the exogeneity of our instruments, we need to ensure that changes in local wind direction affect student achievement only through their effect on the air pollution exposure of the student on the day of the exam. This assumption is inherently untestable but our empirical specification rules out other potential channels through which changes in local wind direction can plausibly impact student performance. For instance, by controlling for weather conditions we rule out other weather phenomenon that are correlated with changes in local wind directions and student performance.

Our second stage specification is as follows.

$$y_{is} = \alpha_i + \rho_s + \mathrm{W}_{e(is(t))}\gamma + \beta \widehat{PM2.5}_{e(is(t))} + \pi_{is} \tag{24}$$

where $\widehat{PM2.5}_{e(is(t))}$ is the predicted $PM2.5$ concentration from the first stage. Other parameters in equation (24) are same as in equation (22).

In Figure E2, we show how the direction from which the wind is blowing affects the $PM2.5$ concentration at examination centers during exam days in our sample. As we expect the wind blowing from the oceans to be relatively cleaner, wind from the Indian Ocean increases pollutants relatively less than winds from the land. We also show in Figure E3 how wind direction using 10 degree bins affects air pollution concentration for each of the nine examination center clusters. We conclude that the same wind direction exerts a differential impact on $PM2.5$ concentration for different examination center clusters.

3.5 Results

3.5.1 Main Results

We present point estimates from OLS and IV specifications in Table 14. In the first three columns, we use as dependent variable aggregate score of the student. For illustrative purposes, we present results from a specification in which we do not include any fixed effects in the first column. Our OLS results, while getting attenuated after including fixed effects, show an increase in $PM2.5$ concentration by $10\ \mu g/m^3$ on days during which the student appears for an exam, increasing their aggregate score by 0.02 standard deviations, on average. To overcome issues associated with classical measurement error in air pollution exposure and time-varying unobservable characteristics that affect both air pollution and student achievement, we report results from the IV specification in the third column. We conclude that an increase in $PM2.5$ concentration by $10\ \mu g/m^3$ on days during which the student appears for an exam increases their aggregate score by 0.37 standard deviations, on average.[4] Next, we present results from estimating equation (22) in the three following columns. Our preferred IV specification in the last column implies that an increase in $PM2.5$ concentration exposure of a student by $10\ \mu g/m^3$ on the day they appear for their exam increases their score on the exam by 0.05 standard deviations, on average. Although smaller, the estimation of OLS specification indicates a similar worsening of student achievement due to higher air pollution exposure on the examination day. We discuss how these point estimates compare to earlier work in section 3.6.

We also examine dose response of air pollution exposure on student achievement. To this end, we replace the $PM2.5$ concentration measure with an indicator for the $PM2.5$ concentration to be greater than $15\ \mu g/m^3$. This threshold correspond to the World Health Organization's safe $PM2.5$ concentration level.[5] We present results from this exercise in Table 15. Our point estimates indicate that extremely polluted days on which the exams take place lead to relatively larger decline in student performance on the exam.

Finally, we also look at the extensive margin effect of how air pollution exposure affects whether the student appears for the exam or not. In addition, we also examine how air pollution exposure on the day of exam impacts whether student passes the exam or not. For the former, we use our main specifications with the dependent variable being an indicator for whether the student appeared or not for exam in the subject for which they registered to take exam. While for the

[4]Note that first stage in both IV specifications is above the threshold beyond which instruments are considered to be strong (Lee et al., 2021).

[5]More information on WHO threshold is available at https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health

Table 14. Effect of $PM2.5$ Concentration on Student Achievement – Main Effect

	(1) OLS	(2) OLS	(3) IV	(4) OLS	(5) OLS	(6) IV
$PM2.5\ (\mu g/m^3)$	0.00096	0.00212	0.03773	0.00011	0.00115	0.00525
	(0.00074)	(0.00067)***	(0.00317)***	(0.00042)	(0.00011)***	(0.00128)***
Weather Controls	✓	✓	✓	✓	✓	✓
Student FE					✓	✓
Subject FE					✓	✓
Exam Level FE		✓	✓			
Examination Centre FE		✓	✓			
Year FE		✓	✓			
KP F-Statistic			208.727			161.365
Examination Centres	6049	6049	6049	6058	6058	6058
N	1,787,993	1,787,993	1,787,993	13,916,070	13,916,070	13,916,070

Notes: Heteroskedasticity robust standard errors clustered by the examination centre are in parentheses. (* p<.10 ** p<.05 *** p<.01). Each observation in column (1) to column (3) corresponds to a unique student. Each observation in column (4) to column (6) corresponds to a unique subject and is at the student level. The dependent variable in column (1) to column (3) is the standardized aggregate score of the student in either the CSEE or ACSEE exam. The dependent variable in column (4) to column (6) is the standardized student score in a subject for which they appeared. Each column regresses the dependent variable on the $PM2.5$ concentration measure, weather controls, and set of fixed effects. Weather controls contain cumulative precipitation, temperature, dew point temperature, sea level pressure, and wind speed in flexible non-parametric form, with the indicator for the observation to be in one of the quartiles of the overall distribution over the sample period. The sample contains data on exams conducted from 2016 to 2019.

latter, we use an indicator for whether the student passes the exam or not for which they appeared as a dependent variable. We present results from this analysis in Table 16. Our point estimates indicate that there is no effect on the extensive margin of whether the student appear for exams or not. However, we find that elevated levels of air pollution on the day of exam increases the likelihood that student failed the exam for which they appeared. Together, these results suggest two implications. First, there is no change in the composition of the students who appear for exam on days of high air pollution. Second, the elevated air pollution levels manifest in an extreme outcome where the student fails to pass the exam. Results in this section allow us to conclude that air pollution worsens student achievement. We explore how this effect varies across subpopulations in section 3.5.3.

Table 15. Effect of $PM2.5$ Concentration on Student Achievement – Dose Response

	(1) OLS	(2) OLS	(3) IV	(4) IV	(5) OLS	(6) OLS	(7) IV	(8) IV
$PM2.5$ ($\mu g/m^3$)	0.00212		0.03774		0.00115		0.00525	
	(0.00067)***		(0.00317)***		(0.00011)***		(0.00128)***	
$PM2.5$ ($> 15\mu g/m^3$)=1		0.09498		0.71773		0.02543		0.20068
		(0.01561)***		(0.06405)***		(0.00316)***		(0.03015)***
Weather Controls	✓	✓	✓	✓	✓	✓	✓	✓
Student FE					✓	✓	✓	✓
Subject FE					✓	✓	✓	✓
Exam Level FE	✓	✓	✓	✓				
Examination Centre FE	✓	✓	✓	✓				
Year FE	✓	✓	✓	✓				
KP F-Statistic			208.712	238.368			161.365	26.362
Examination Centres	6049	6049	6049	6049	6058	6058	6058	6058
N	1,787,993	1,787,993	1,787,993	1,787,993	13,916,070	13,916,070	13,916,070	13,916,070

Notes: Heteroskedasticity robust standard errors clustered by the examination centre are in parentheses. (* p<.10 ** p<.05 *** p<.01). Each observation in column (1) to column (4) corresponds to a unique student. Each observation in column (5) to column (8) corresponds to a unique subject and is at the student level. The dependent variable in column (1) to column (4) is the standardized aggregate score of the student in either the CSEE or ACSEE exam. The dependent variable in column (5) to column (8) is the standardized student score in a subject for which they appeared. Each column regresses the dependent variable on the $PM2.5$ concentration measure, weather controls, and set of fixed effects. Weather controls contain cumulative precipitation, temperature, dew point temperature, sea level pressure, and wind speed in flexible non-parametric form, with the indicator for the observation to be in one of the quartiles of the overall distribution over the sample period. The sample contains data on exams conducted from 2016 to 2019.

3.5.2 Robustness Checks

In order to establish the robustness of our results, we conduct a few exercises. We present results from these in Table 17. First, we include a lead of the $PM2.5$ measure in our baseline specifications from equation (22). With this exercise, we aim to assuage concerns regarding contemporaneous events that might be correlated with air pollution and affect student exam performance. Our results indicate that this concern is not valid in our setting. Our point estimates are not altered when we include a lead of the $PM2.5$ concentration in our main specification. To address concerns about our effects being convoluted by other co-pollutants, we augment our baseline specifications with a measure of Ozone (O_3) on the day of the exam. Our results indicate that even after including a concentration measure for Ozone, our point estimates are similar to the

Table 16. Effect of $PM2.5$ Concentration on Student Achievement – Extensive Margin Effect

	(1) OLS	(2) OLS	(3) OLS	(4) OLS
$PM2.5\ (\mu g/m^3)$	0.00001	0.00022	0.00000	0.00068
	(0.00013)	(0.00020)	(0.00001)	(0.00005)***
Weather Controls	✓	✓	✓	✓
Student FE			✓	✓
Subject FE			✓	✓
Exam Level FE	✓	✓		
Examination Centre FE	✓	✓		
Year FE	✓	✓		
Outcome Mean	0.07712	0.19353	0.02646	0.40803
Examination Centres	6060	6060	6060	6060
N	1,937,410	1,937,410	14,294,285	14,294,285

Notes: Heteroskedasticity robust standard errors clustered by the examination centre are in parentheses. (* p<.10 ** p<.05 *** p<.01). Each observation in column (1) and column (2) corresponds to a unique student. Each observation in column (3) and column (4) corresponds to a unique subject and is at the student level. The dependent variable in column (1) and column (3) is an indicator for whether the student appeared for the exams that they registered for. The dependent variable in column (2) and column (4) is an indicator for whether the student failed the exam that they appeared for. Each column regresses the dependent variable on the $PM2.5$ concentration measure, weather controls, and set of fixed effects. Weather controls contain cumulative precipitation, temperature, dew point temperature, sea level pressure, and wind speed in flexible non-parametric form, with the indicator for the observation to be in one of the quartiles of the overall distribution over the sample period. The sample contains data on exams conducted from 2016 to 2019.

main results in section 3.5.1. We also show the robustness of our results by restricting the sample to those subjects that did not have exams spread over multiple days. We present results from this exercise in the second last column of Table 17. We find that our point estimate is slightly higher than the point estimate from our main results and continues to be highly statistically significant.

Finally, we examine if our results are robust to excluding students who reappear for high-stakes exams. To this end, we use information on whether the candidate appears for exams at a private examination center or a school. Since examination rules require reappearing candidates to appear only at a private examination center, the robustness of our main results to excluding candidates that appear for exams at such centers will provide support for the interpretation that our results are not driven by candidates who reappear for these high-stakes exams. We present results from this exercise in the last column of Table 17. Our results suggest that when we exclude reappearing candidates, the point estimate is slightly lower and continues to be highly statistically significant. To conclude, these exercises help us establish that it is contemporaneous $PM2.5$ concentration on the day of the exam that impacts student performance (Bedi et al., 2021; Carneiro et al., 2021).

Table 17. Effect of $PM2.5$ Concentration on Student Achievement – Robustness Check

	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS	(8) OLS
$PM2.5\ (\mu g/m^3)$	0.00212	-0.00028	0.00232	0.00115	0.00115	0.00088	0.00140	0.00109
	(0.00067)***	(0.00233)	(0.00068)***	(0.00011)***	(0.00012)***	(0.00011)***	(0.00013)***	(0.00012)***
$PM2.5\,(1\ Day\ Lead)\ (\mu g/m^3)$		0.00146			-0.00010			
		(0.00187)			(0.00011)			
$O_3\ (kg/m^2)$			394.89496			384.23433		
			(42.18261)***			(17.56887)***		
Weather Controls	✓	✓	✓	✓	✓	✓	✓	✓
Student FE				✓	✓	✓	✓	✓
Subject FE				✓	✓	✓	✓	✓
Exam Level FE	✓	✓	✓					
Examination Centre FE	✓	✓	✓					
Year FE	✓	✓	✓					
Examination Centres	6049	6049	6049	6058	6058	6058	6058	4897
N	1,787,993	1,787,993	1,787,993	13,916,070	13,916,070	13,916,070	10,048,995	12,610,795

Notes: Heteroskedasticity robust standard errors clustered by the examination centre are in parentheses. (* p<.10 ** p<.05 *** p<.01). Each observation in first three column corresponds to a unique student. Each observation in last three column corresponds to a unique subject and is at the student level. The dependent variable in first three column is the standardized aggregate score of the student in either the CSEE or ACSEE exam. The dependent variable in last five column is the standardized student score in a subject for which they appeared. Each column regresses the dependent variable on the $PM2.5$ concentration measure, weather controls, and set of fixed effects. Weather controls contain cumulative precipitation, temperature, dew point temperature, sea level pressure, and wind speed in flexible non-parametric form, with the indicator for the observation to be in one of the quartiles of the overall distribution over the sample period. The sample contains data on exams conducted from 2016 to 2019.

3.5.3 Heterogeneity Analysis

In this section, we analyze how worse air quality exerts a differential impact on various subpopulations. To this end, we augment specification in equation (22) by interacting our main regressor, $PM2.5$ concentration, with indicators for different subpopulations. In particular, we explore heterogeneous effects of air pollution by sex of the student, age of the student, whether the student appeared for the exam in an examination center that is private owned or government owned, where the student is in the exam score distribution, and by different types of subjects. We discuss each of these below.

We present results from estimating the specification in equation (22) in Table 18. The first column in the table repeat results from Table 14. We find that female students suffer a more pronounced worsening of exam performance relative to male students due to exposure to elevated levels of air pollution on examination days. While this result is contrary to what some of the earlier work has found, we believe contextual factors might be responsible for this finding (Carneiro et al., 2021; Ebenstein et al., 2016). For instance, some of the previous work has either looked at urban areas or developed countries where gender bias in access to resources might not exist. In our setting, multiple previous studies show that there is gender bias against girls in access to resources like nutrition, health, and education both during childhood and adulthood (Al-Samarrai and Peasgood, 1998; Dercon and Krishnan, 2000; Hadley et al., 2008; Klasen, 1996). To the extent that this reduces the strength of females in coping with elevated levels of air pollution, we expect to see relatively worse outcomes for females than for males (Balakrishnan and Tsaneva, 2021; Balietti et al., 2022; Ridolo et al., 2019).

We examine how the student's age affects the relationship between air pollution exposure and exam performance. We augment our specifications by interacting the measure of $PM2.5$ concentration with an indicator for whether the student appeared for the CSEE or ACSEE exam. As we discussed in section 3.3.1, students take the CSEE exam in Grade 10 while the ACSEE is administered in Grade 12. Results from our preferred specification in equation (22) show that the detrimental impact of air pollution exposure on exam performance is more pronounced for younger students. This finding is similar to what previous work elsewhere has found (Austin et al., 2019; Chen et al., 2018; Currie et al., 2009; Singh, 2022). Next, we look at whether the type of examination center where students appear for the exam imposes a differential impact on how air pollution affects student achievement. Similar to the exercises above, we augment our baseline specification by interacting with our main regressor with whether the examination center is private or government owned. Our results show that the effect of air pollution in worsening

Table 18. Effect of $PM2.5$ Concentration on Student Achievement – Heterogeneous Effect

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	OLS	OLS	OLS	OLS	OLS	OLS	OLS
$PM2.5\ (\mu g/m^3)$	0.00115	0.00355	-0.01194	0.00145	0.01153	-0.01038	0.00117
	(0.00011)***	(0.00016)***	(0.00266)***	(0.00013)***	(0.00031)***	(0.00037)***	(0.00013)***
Male × $PM2.5\ (\mu g/m^3)$		-0.00513					
		(0.00020)***					
CSEE × $PM2.5\ (\mu g/m^3)$			0.01321				
			(0.00266)***				
Private × $PM2.5\ (\mu g/m^3)$				-0.00169			
				(0.00029)***			
Score Below Median × $PM2.5\ (\mu g/m^3)$					-0.04239		
					(0.00084)***		
Score Bottom 5 Percentile × $PM2.5\ (\mu g/m^3)$						-0.02065	
						(0.00063)***	
Score Top 5 Percentile × $PM2.5\ (\mu g/m^3)$						0.03390	
						(0.00073)***	
Real GDP Below Median × $PM2.5\ (\mu g/m^3)$							0.00284
							(0.00041)***
Weather Controls	✓	✓	✓	✓	✓	✓	✓
Student FE	✓	✓	✓	✓	✓	✓	✓
Subject FE	✓	✓	✓	✓	✓	✓	✓
Examination Centres	6058	6058	6058	4497	6058	6058	5385
N	13,916,070	13,916,070	13,916,070	11,674,613	13,916,070	13,916,070	12,083,006

Notes: Heteroskedasticity robust standard errors clustered by the examination centre are in parentheses. (* p<.10 ** p<.05 *** p<.01). Each observation in column (1) to column (7) corresponds to a unique subject and is at the student level. The dependent variable in each column is the standardized student score in a subject for which they appeared. Each column regresses the dependent variable on the $PM2.5$ concentration measure, weather controls, and set of fixed effects. Weather controls contain cumulative precipitation, temperature, dew point temperature, sea level pressure, and wind speed in flexible non-parametric form, with the indicator for the observation to be in one of the quartiles of the overall distribution over the sample period. The sample contains data on exams conducted from 2016 to 2019.

student achievement is more pronounced for those who appear for exams at government-owned examination centers.

We also look at how students at different levels in the achievement distribution are impacted by air pollution. We undertake two exercises to do this. First, we interact the $PM2.5$ concentration measure in our baseline specification with whether the student scored below or above the median in their examination center and subject group. Lastly, we interact the air pollution exposure measure with a categorical variable with three levels – whether the student scored in the top five percentile, whether the student scored in the bottom five percentile, and whether the student scored in neither the top five nor the bottom five percentile of exam score in their examination center and subject group. Results from these exercises indicate that weaker students see a relatively greater decline in their exam performance. This finding is similar to what previous work has established (Carneiro et al., 2021).

We next examine how air pollution exposure on the day of the exam affects student performance on the exam, depending on whether they belong to relatively poorer regions of the country. To this end, we obtain information on the real gross domestic product (GDP) per capita for 2015 from the Tanzania National Bureau of Statistics. We classify regions as being above or below the median real GDP per capita. We then interact this indicator variable with our $PM2.5$ concentration measure in equation 22. We present results from this exercise in the seventh column of Table 18. We conclude that students who appear for exams at examination centers in relatively poorer regions see a disproportionately larger decline in performance. This finding suggests that inequalities in air pollution exposure on the day of the exam may further exacerbate the existing achievement gap (Akmal and Pritchett, 2021; Chmielewski, 2019).

Next, we examine how air pollution impacts students' performance in various subjects. We present results from these exercises in Table 19. We find that relative to non-reading and non-quantitative subjects, the adverse impact on exam performance is more pronounced for subjects that are quantitative in nature compared to reading subjects if students are exposed to elevated levels of air pollution on the day of the exam. The difference in point estimates for these two sets of subjects is statistically significant.[6] This result is what previous work has documented (Bedi et al., 2021; La Nauze and Severnini, 2021). Our large sample size allows us to precisely estimate these effects lending further support to the findings that adverse effects on fluid intelligence might be the mechanism through which air pollution impacts cognitive outcomes (Loftus et al., 2019). Fluid intelligence is the ability to solve flexible and novel problems. To further examine the heterogeneity between subjects that test fluid and non-fluid intelligence, in the last column of the table, we categorize subjects into these two categories. Our point estimates show that the effect is more pronounced for subjects that test fluid intelligence.

[6] F-statistic for a test of point estimate for quantitative subjects being equal to point estimate for reading subjects is 4.75 with a p-value of 0.0293. Therefore, we reject the null hypothesis that these two point estimates are not statistically different from each other.

Table 19. Effect of $PM2.5$ Concentration on Student Achievement – Heterogeneous Effect

	(1) OLS	(2) OLS	(3) OLS
$PM2.5\ (\mu g/m^3)$	0.00115	-0.00001	0.00115
	(0.00011)***	(0.00014)	(0.00011)***
Quantitative Subjects × $PM2.5\ (\mu g/m^3)$		0.00199	
		(0.00019)***	
Reading Subjects × $PM2.5\ (\mu g/m^3)$		0.00153	
		(0.00014)***	
Fluid Reasoning Subjects × $PM2.5\ (\mu g/m^3)$			0.00588
			(0.00243)**
Weather Controls	✓	✓	✓
Student FE	✓	✓	✓
Subject FE	✓	✓	✓
Examination Centres	6058	6058	6058
N	13,916,070	13,916,070	13,916,070

Notes: Heteroskedasticity robust standard errors clustered by the examination centre are in parentheses. (* p<.10 ** p<.05 *** p<.01). Each observation in column (1) to column (3) corresponds to a unique subject and is at the student level. The dependent variable in each column is the standardized student score in a subject for which they appeared. Each column regresses the dependent variable on the $PM2.5$ concentration measure, weather controls, and set of fixed effects. Weather controls contain cumulative precipitation, temperature, dew point temperature, sea level pressure, and wind speed in flexible non-parametric form, with the indicator for the observation to be in one of the quartiles of the overall distribution over the sample period. Fluid Subjects in the third column are architectural draughting, brickwork and masonry, building construction, carpentry, electrical draughting, electrical installation, fitting and turning, mechanical draughting, motor vehicle mechanics, painting and signwriting, plant and equipment maintenance, plumbing, radio and tv, and refrigeration and air conditioning. The sample contains data on exams conducted from 2016 to 2019.

3.6 Discussion and Conclusion

In this paper, we study how air pollution on the day of the exam affects student achievement on the exam. We utilize novel data on student exam scores on national examinations during secondary schooling in Tanzania. We leverage plausibly exogenous changes in local wind direction in an instrumental variables setup to uncover causal effects. We find that higher levels of air pollution on the day of the exam worsen student achievement. Our heterogeneity analyses reveal that this effect is more pronounced for younger students, females, students who appear for the exam in examination centers that are located in government schools, students in poorer regions, and students who are at the lower end of the achievement distribution. Further, we leverage detailed information on the student scores on various exams for which they appeared to provide suggestive evidence that the main effect of air pollution on student achievement is driven by worse performance on high air pollution days in exams that test fluid intelligence like quantitative subjects. We also establish the robustness of our results to multiple falsification checks.

Our point estimates are comparable to the effect of air pollution on student achievement documented before in the context of developed countries or small geographical areas. Specifically, our IV estimates imply that an increase in $PM2.5$ concentration by $10\ \mu g/m^3$ on the day a student appears for the exam increases their score on the exam by 0.05 standard deviations, thus worsening their performance on these exams. In order to benchmark this effect, we refer to a large literature on the effect of teacher value-added on student achievement. Bau and Das (2020) show that an increase in 1 standard deviation in teacher value-added improves student test scores by 0.15 standard deviations in a low-income country. Extrapolating this effect size to our context, we find that an increase in $PM2.5$ concentration by $10\ \mu g/m^3$ on the day a student appears for the exam is approximately equivalent to a worsening of teacher value-added by approximately 0.33 standard deviations.

While we are confident that we document the causal effect of air pollution on student achievement, our work has several limitations. First, data limitation does not allow us to identify what sections of the exam drive the worsening in academic performance when students are exposed to higher levels of air pollution. Second, although we show that the main effect is driven by a potential worsening of performance in tests of fluid intelligence, our evidence is indirect and suggestive at best. Third, we are limited by data constraints in being able to study only the contemporaneous effects of air pollution on student achievement. Finally, we are unable to study how air pollution on the day of exam evaluation affects student performance as evaluated by examiners. We hope future work is able to answer some of these important questions.

Appendix A. Mathematical Proofs for Chapter I

A.1 Proof of Lemma 1

Proof. OLS estimation biasdness: Least squares estimates of the parameters, $\theta_{LS} = [\alpha_1,\ \alpha_2,\ \alpha_3,\ \beta]'$, in the operation model in equation (3) is given by

$$\hat{\theta}_{LS} = (Z'Z)^{-1}Z'y$$

where $Z = [t_1,\ t_2,\ t_3,\ x]$ and $t_3 = t_1 \times t_2$. It follows that

$$\hat{\theta}_{LS} = (Z'Z)^{-1}Z'(Z^{*}\theta + \epsilon)$$

such that $Z = [t_1^{*},\ t_2^{*},\ t_3^{*},\ x]$. $Z'Z$ is positive-semi definite by Cauchy-Schwarz inequality. Therefore, the bias term in OLS bias, given by $\hat{\theta}_{LS} = (Z'Z)^{-1}Z'Z^{*}\theta + (Z'Z)^{-1}Z'\epsilon$, is determined by $Z'Z^{*}$ and $Z'\epsilon$.

OLS estimation inconsistency: We can express $\hat{\theta}_{LS} = (Z'Z)^{-1}Z'Z^{*}\theta + (Z'Z)^{-1}Z'\epsilon$ as

$$\hat{\theta}_{LS} = \left(\frac{Z'Z}{n}\right)^{-1}\left(\frac{Z'Z^{*}}{n}\right)\theta + \left(\frac{Z'Z}{n}\right)^{-1}\frac{Z'\epsilon}{n}$$

By taking the probability limit and applying Slutsky's Lemma, we have

$$\text{plim}\ \hat{\theta}_{LS} = \text{plim}\left(\frac{Z'Z}{n}\right)^{-1}\text{plim}\left(\frac{Z'Z^{*}}{n}\right)\theta + \text{plim}\left(\frac{Z'Z}{n}\right)^{-1}\text{plim}\frac{Z'\epsilon}{n}$$

By the Weak Law of Large Numbers and continuous mapping theorem, we have

$$\text{plim}\left(\frac{Z'Z}{n}\right)^{-1} \xrightarrow{p} (\text{Cov}\,(Z,Z))^{-1}$$

Following similar argument as above, we have

$$\text{plim}\frac{Z'Z^{*}}{n} \xrightarrow{p} \text{Cov}\,(Z,Z^{*})$$

and by exogeneity of covariates in Z, we have

$$\text{plim}\frac{Z'\epsilon}{n}\xrightarrow{p}\text{Cov}\left(Z,\epsilon\right)=0$$

Combining the expressions above, we can easily obtain the results in Lemma 1, that is

$$\text{plim }\hat{\theta}_{LS}=\left(\text{Cov}\left(Z,Z\right)\right)^{-1}\text{Cov}\left(Z,Z^{*}\right)\theta$$

Proof. Joint treatment effect OLS estimation biasdness: The least squares estimates of α_3 is given by

$$\hat{\alpha}_{3LS}=\left(t_3'M_{-t_3}t_3\right)^{-1}t_3M_{-t_3}y$$

Where $M_{-t_3}=1-Z_{-t_3}\left(Z'_{-t_3}Z_{-t_3}\right)^{-1}Z'_{-t_3}$ and $Z_{-t_3}=[t_1,\ t_2,\ x]$. It follows that

$$\hat{\alpha}_{3LS}=\left(t_3'M_{-t_3}t_3\right)^{-1}t_3M_{-t_3}\left(Z^{*}_{-t_3}\theta_{-\alpha_3}+\alpha_3t_3^{*}+\epsilon\right)$$

and

$$\hat{\alpha}_{3LS}-\alpha_3=\left(t_3'M_{-t_3}t_3\right)^{-1}t_3M_{-t_3}\left(Z^{*}_{-t_3}\theta_{-\alpha_3}+\alpha_3\left(t_3^{*}-t_3\right)+\epsilon\right)$$

By Cauchy-Schwarz inequality, $\left(t_3'M_{-t_3}t_3\right)^{-1}$ is positive semi-definite. Hence, the bias in OLS estimates, given by $\hat{\alpha}_{3LS}-\alpha_3=\left(t_3'M_{-t_3}t_3\right)^{-1}\left(t_3M_{-t_3}Z^{*}_{-t_3}\theta_{-\alpha_3}+t_3M_{-t_3}\left(t_3^{*}-t_3\right)\alpha_3+t_3M_{-t_3}\epsilon\right)$, is driven by the remaining terms, that is, $t_3M_{-t_3}Z^{*}_{-t_3}$, $t_3M_{-t_3}t_3^{*}$, and $t_3M_{-t_3}\epsilon$.

Joint treatment effect OLS estimation inconsistency: We can express the bias term in least squares estimator as

$$\hat{\alpha}_{3LS}-\alpha_3=\left(\frac{t_3'M_{-t_3}t_3}{n}\right)^{-1}\left(\frac{t_3M_{-t_3}Z^{*}_{-t_3}}{n}\theta_{-\alpha_3}+\frac{t_3M_{-t_3}\left(t_3^{*}-t_3\right)}{n}\alpha_3+\frac{t_3M_{-t_3}\epsilon}{n}\right)$$

Taking probability limit and using Slutsky Lemma, we have

$$\text{plim } \hat{\alpha}_{3LS} - \alpha_3 = \text{plim}\left(\frac{t_3' M_{-t_3} t_3}{n}\right)^{-1}\left(\text{plim}\frac{t_3 M_{-t_3} Z_{-t_3}^*}{n}\theta_{-\alpha_3} + \text{plim}\frac{t_3 M_{-t_3}\left(t_3^* - t_3\right)}{n}\alpha_3 + \text{plim}\frac{t_3 M_{-t_3}\epsilon}{n}\right)$$

Expanding the projection matrix M_{-t_3}, we can then express plim $\frac{t_3' M_{-t_3} t_3}{n}$ as

$$\begin{aligned}\text{plim } \frac{t_3' M_{-t_3} t_3}{n} &= \text{plim } \frac{t_3'\left(1 - Z_{-t_3}\left(Z_{-t_3}' Z_{-t_3}\right)^{-1} Z_{-t_3}'\right) t_3}{n} \\ &= \text{plim } \frac{t_3' t_3}{n} - \text{plim } \frac{t_3' Z_{-t_3}\left(Z_{-t_3}' Z_{-t_3}\right)^{-1} Z_{-t_3}' t_3}{n}\end{aligned}$$

By Weak Law of Large Numbers, we have

$$\text{plim } \frac{t_3' M_{-t_3} t_3}{n} = \mathbb{E}\left[t_{3i}\right] - \mathbb{E}\left[t_{3i}' \mathrm{z}_{i,-t_{3i}}\right]\left(\mathbb{E}\left[\mathrm{z}_{i,-t_{3i}}' \mathrm{z}_{i,-t_{3i}}\right]\right)^{-1} \mathbb{E}\left[\mathrm{z}_{i,-t_{3i}} t_{3i}\right]$$

By continuous mapping theorem, we have

$$\left(\text{plim } \frac{t_3' M_{-t_3} t_3}{n}\right)^{-1} = \left(\mathbb{E}\left[t_{3i}\right] - \mathbb{E}\left[t_{3i}' \mathrm{z}_{i,-t_{3i}}\right]\left(\mathbb{E}\left[\mathrm{z}_{i,-t_{3i}}' \mathrm{z}_{i,-t_{3i}}\right]\right)^{-1} \mathbb{E}\left[\mathrm{z}_{i,-t_{3i}} t_{3i}\right]\right)^{-1}$$

Applying the arguments above to plim $\frac{t_3 M_{-t_3} Z_{-t_3}^*}{n}\theta_{-\alpha_3}$ and plim $\frac{t_3 M_{-t_3}\left(t_3^* - t_3\right)}{n}$, It follows that

$$\begin{aligned}\text{plim } \frac{t_3 M_{-t_3} Z_{-t_3}^*}{n}\theta_{-\alpha_3} &= \mathbb{E}\left[t_{3i} \mathrm{z}_{i,-t_{3i}}^{\prime *}\theta_{-\alpha_3}\right] \\ &\quad - \mathbb{E}\left[t_{3i}\mathrm{z}_{i,-t_{3i}}'\right]\left(\mathbb{E}\left[\mathrm{z}_{i,-t_{3i}}\mathrm{z}_{i,-t_{3i}}'\right]\right)^{-1}\mathbb{E}\left[\mathrm{z}_{i,-t_{3i}}\mathrm{z}_{i,-t_{3i}}^*\theta_{-\alpha_3}\right]\end{aligned}$$

and

$$\text{plim}\ \frac{t_3 M_{-t_3}\left(t_3^* - t_3\right)}{n} = \mathbb{E}\left[t_{3i}\left(t_{3i}^* - t_{3i}\right)\right]$$
$$- \mathbb{E}\left[t_{3i}\mathrm{z}'_{i,-t_{3i}}\right]\left(\mathbb{E}\left[\mathrm{z}_{i,-t_{3i}}\mathrm{z}'_{i,-t_{3i}}\right]\right)^{-1}\mathbb{E}\left[\mathrm{z}_{i,-t_{3i}}\left(t_{3i}^* - t_{3i}\right)\right]$$
$$= -\mathbb{E}\left[t_{3i}\mathrm{z}'_{i,-t_{3i}}\right]\left(\mathbb{E}\left[\mathrm{z}_{i,-t_{3i}}\mathrm{z}'_{i,-t_{3i}}\right]\right)^{-1}\mathbb{E}\left[\mathrm{z}_{i,-t_{3i}}\left(t_{3i}^* - t_{3i}\right)\right]$$

Finally, following the exogeneity of t_3 and other covariates in the model, we have

$$\text{plim}\ \frac{t_3 M_{-t_3}\epsilon}{n} = \mathbb{E}\left[t_{3i}\epsilon'_i\right] - \mathbb{E}\left[t_{3i}\mathrm{z}'_{i,-t_{3i}}\right]\left(\mathbb{E}\left[\mathrm{z}_{i,-t_{3i}}\mathrm{z}'_{i,-t_{3i}}\right]\right)^{-1}\mathbb{E}\left[\mathrm{z}_{i,-t_{3i}}\epsilon_i\right] = 0$$

We get the desired results, the asymptotic bias of the joint treatment effect OLS estimator, α_{3LS}, by combining all the terms above, that is

$$\text{plim}\quad \hat{\alpha}_{3LS} - \alpha_3 = \frac{A - B\alpha_3}{Q}$$

Where $A = \mathbb{E}\left[t_{3i}\mathrm{z}'^{*}_{i,-t_{3i}}\theta_{-\alpha_3}\right] - \mathbb{E}\left[t_{3i}\mathrm{z}'_{i,-t_{3i}}\right]\left(\mathbb{E}\left[\mathrm{z}_{i,-t_{3i}}\mathrm{z}'_{i,-t_{3i}}\right]\right)^{-1}\mathbb{E}\left[\mathrm{z}_{i,-t_{3i}}\mathrm{z}^*_{i,-t_{3i}}\theta_{-\alpha_3}\right]$,

$B = \mathbb{E}\left[t_{3i}\mathrm{z}'_{i,-t_{3i}}\right]\left(\mathbb{E}\left[\mathrm{z}_{i,-t_{3i}}\mathrm{z}'_{i,-t_{3i}}\right]\right)^{-1}\mathbb{E}\left[\mathrm{z}_{i,-t_{3i}}\left(t_{3i}^* - t_{3i}\right)\right]$,

$Q = \mathbb{E}\left[t_{3i}\right] - \mathbb{E}\left[t_{3i}\mathrm{z}'_{i,-t_{3i}}\right]\left(\mathbb{E}\left[\mathrm{z}_{i,-t_{3i}}\mathrm{z}'_{-3i}\right]\right)^{-1}\mathbb{E}\left[\mathrm{z}_{i,-t_{3i}}t_{3i}\right]$,

$\theta_{-\alpha_3} = [\alpha_1,\ \alpha_2,\ \beta]'$, and $\mathrm{z}_{i,-t_{3i}} = [t_{1i}; t_{2i}; x_i]$

A.2 Proof of Lemma 2

Proof. Components of estimator of W, $\hat{W}(a_1, a_2)$: W, variance-covariance matrix $\text{Cov}(Z, Z^*)$, is given by

$$W = \begin{bmatrix} \text{Cov}(t_{1i}, t^*_{1i}) & \text{Cov}(t_{1i}, t^*_{2i}) & \text{Cov}(t_{1i}, t^*_{3i}) & \text{Cov}(t_{1i}, x_i) \\ \text{Cov}(t_{2i}, t^*_{1i}) & \text{Cov}(t_{2i}, t^*_{2i}) & \text{Cov}(t_{2i}, t^*_{3i}) & \text{Cov}(t_{2i}, x_i) \\ \text{Cov}(t_{3i}, t^*_{1i}) & \text{Cov}(t_{3i}, t^*_{2i}) & \text{Cov}(t_{3i}, t^*_{3i}) & \text{Cov}(t_{3i}, x_i) \\ \text{Cov}(x_i, t^*_{1i}) & \text{Cov}(x_i, t^*_{2i}) & \text{Cov}(x_i, t^*_{3i}) & \text{Var}(x_i) \end{bmatrix}$$

Given the data, all terms in W except those in the last column, are not directly observed by the researcher since they are determined by true (and unobserved) participation status. Given the misclassification probabilities, a_1 and a_2, the probability of false negative in t_{1i} and t_{2i}, we can obtain the probability of false negative in the interaction term, t_{3i}, as follows.

By the Law of Iterated Expectations, $\mathbb{E}[t^*_{3i}] = \mathbb{E}[t^*_{1i}t^*_{2i}] = \mathbb{E}[\mathbb{E}[t^*_{1i}|t^*_{2i}]\ \mathbb{E}[t^*_{2i}|t^*_{1i}]]$. Considering that $\Pr(t_{ji} = 1) = (1 - a_j)\Pr(t^*_{ji} = 1)$ for $j \in \{1, 2\}$, it follows that $\mathbb{E}[t^*_{3i}] = \mathbb{E}\left[\frac{\mathbb{E}[t_{1i}|t^*_{2i}]}{1 - a_1}\ \frac{\mathbb{E}[t_{2i}|t^*_{1i}]}{1 - a_2}\right] = \frac{\mathbb{E}\,[t_{1i}\,t_{2i}]}{(1 - a_1)(1 - a_{2i})} = \frac{\mathbb{E}\,[t_{3i}]}{1 - a_3}$, where a_3 is the probability of false negative in t_{3i} and is given by $a_3 = a_1 + a_2 - a_1 \times a_2$.

Let's first consider the covariance between observed (and plausibly error-driven) program participation status and the underlying (true) status, that is, $\text{Cov}(\text{t}_{\text{ji}}, \text{t}^*_{\text{ji}})$, for $j \in \{1, 2, 3\}$. We can write

$$\begin{aligned}
\text{Cov}(t_{ji}, t^*_{ji}) &= \mathbb{E}\left[t_{ji}, t^*_{ji}\right] - \mathbb{E}\left[t_{ji}\right]\mathbb{E}\left[t^*_{ji}\right] \\
&= \Pr\left[t_{ji} = 1, t^*_{ji} = 1\right] - \Pr\left[t_{ji} = 1\right]\Pr\left[t^*_{ji} = 1\right] \\
&= \Pr\left[t_{ji} = 1|t^*_{ji} = 1\right]\Pr\left[t^*_{ji} = 1\right] - \Pr\left[t_{ji} = 1\right]\Pr\left[t^*_{ji} = 1\right] \\
&= \left(1 - \Pr\left[t_{ji} = 0|t^*_{ji} = 1\right]\right)\Pr\left[t^*_{ji} = 1\right] - \Pr\left[t_{ji} = 1\right]\Pr\left[t^*_{ji} = 1\right] \\
&= \left(1 - a_j - \Pr\left[t_{ji} = 1\right]\right)\Pr\left[t^*_{ji} = 1\right]
\end{aligned}$$

From $\Pr\left[t_{ji}^* = 1\right] = \left(1 - a_j\right) Pr\left[t_{ji} = 1\right]$, it follows that

$$\begin{aligned}
\text{Cov}(t_{ji}, t_{ji}^*) &= \frac{\left(1 - a_j - \Pr\left[t_{ji} = 1\right]\right) \Pr\left[t_{ji} = 1\right]}{\left(1 - a_j\right)} \\
&= \frac{\left(1 - a_j - \Pr\left[t_{ji} = 1\right]\right) \text{Var}\left(t_{ji}\right)}{\left(1 - a_j\right)\left(\Pr\left[t_{ji} = 1\right]\right)} \\
&= \frac{\left(1 - a_j - P_j\right)}{\left(1 - a_j\right)\left(1 - P_j\right)} \text{Var}\left(t_{ji}\right)
\end{aligned}$$

Hence, for $j \in \{1, 2\}$, $\text{Cov}(t_{ji}, t_{ji}^*)$ can be expressed as

$$\text{Cov}(t_{ji}, t_{ji}^*) = \eta_j \text{Var}\left(t_{ji}\right)$$

where $\eta_j = \dfrac{1 - a_j - P_j}{\left(1 - a_j\right)\left(1 - P_j\right)}$

Next, we examine covariance between true (unobserved) participation status in one program, t_{ji}^* , and the observed participation status in the other program, t_k, that is, $\text{Cov}(\text{t}_{\text{ki}}, \text{t}_{\text{ji}}^*)$ where $j, k \in \{1, 2\}$ and $j \neq k$. By applying the Law of Iterated Expectations, we have

$$\text{Cov}(t_{ki}, t_{ji}^*) = \text{Cov}\left(t_{ki}, \mathbb{E}[t_{ji}^* | t_{ki}]\right)$$

It follows that $\mathbb{E}[t_{ji}^* | t_{ki}] = \dfrac{\mathbb{E}[t_{ji} | t_{ki}]}{1 - a_j}$, which implies that

$$\begin{aligned}
\text{Cov}(t_{ki}, t_{ji}^*) &= \text{Cov}\left(t_{ki}, \frac{\mathbb{E}[t_{ji} | t_{ki}]}{1 - a_j}\right) \\
&= \frac{1}{1 - a_j} \text{Cov}\left(t_{ki}, \mathbb{E}[t_{ji} | t_{ki}]\right) \\
&= \frac{1}{1 - a_j} \text{Cov}\left(t_{ki}, t_{ji}\right)
\end{aligned}$$

We now have the results in Lemma 1, that is, for $j, k \in \{1, 2\}$ and $j \neq k$,

$$\mathrm{Cov}(t_{ki}, t^*_{ji}) = \zeta_j \mathrm{Cov}\left(t_{ki}, t_{ji}\right)$$

where $\zeta_j = \dfrac{1}{1 - a_j}$

We then turn to the covariance between individual program participation and the interaction term given by $\mathrm{Cov}(\mathrm{t_{ji}}, \mathrm{t^*_{3i}})$, for $j \in \{1, 2\}$.

$$\mathrm{Cov}(t_{ji}, t^*_{3i}) = \mathbb{E}\left[t_{ji} t^*_{3i}\right] - \mathbb{E}\left[t_{ji}\right] \mathbb{E}\left[t^*_{3i}\right]$$

We know that $\mathbb{E}\left[t^*_{3i}\right] = \dfrac{\mathbb{E}[t_{3i}]}{1 - a_3}$ and $\mathbb{E}\left[t_{ji} t^*_{3i}\right] = \mathbb{E}\left[t_{ji}\, t^*_{ki}\right]$, so that, by the law of Iterated expectation,

$$\mathbb{E}\left[t_{ji} t^*_{3i}\right] = \mathbb{E}\left[t_{ji} \mathbb{E}[t^*_{ki} | t_{ji}]\right] = \mathbb{E}\left[t_{ji} \frac{\mathbb{E}[t_{ki} | t_{ji}]}{1 - a_k}\right] = \frac{\mathbb{E}\left[t_{ji} t_{ki}\right]}{1 - a_k} = \frac{\mathbb{E}\left[t_{3i}\right]}{1 - a_k}$$

where $k = \{1, 2\}$ and $k \neq j$. Since $t_{3i} = t_{ji} \times t_{ki} = t_{ji} \times t_{3i}$, it follows that

$$\begin{aligned}
\mathrm{Cov}(t_{ji}, t^*_{3i}) &= \frac{\mathbb{E}\left[t_{ji} t_{3ii}\right]}{1 - a_k} - \frac{\mathbb{E}\left[t_{ji}\right] \mathbb{E}\left[t_{3i}\right]}{1 - a_3} \\
&= \frac{\mathbb{E}\left[t_{ji} t_{3i}\right]}{(1 - a_k)} - \frac{\mathbb{E}\left[t_{ji}\right] \mathbb{E}\left[t_{3i}\right]}{(1 - a_j)(1 - a_k)} \\
&= \frac{\mathbb{E}\left[t_{ji} t_{3ii}\right] - \mathbb{E}\left[t_{ji}\right] \mathbb{E}\left[t_{3i}\right] - a_j \mathbb{E}\left[t_{ji} t_{3i}\right]}{(1 - a_j)(1 - a_k)} \\
&= \frac{\mathrm{Cov}\left(t_{ji}, t_{3i}\right) - a_j \mathbb{E}\left[t_{3i}\right]}{(1 - a_j)(1 - a_k)} \\
&= \left[\frac{\mathrm{Cov}\left(t_{ji}, t_{3i}\right) - a_j P_3}{(1 - a_j)(1 - a_k) \mathrm{Cov}\left(t_{ji}, t_{3i}\right)}\right] \mathrm{Cov}\left(t_{ji}, t_{3i}\right)
\end{aligned}$$

We get the results in Lemma 2 if we rewrite the Cov $\left(t_{ji}, t_{3i}\right)$ terms in the bracket as Cov $\left(t_{ji}, t_{3i}\right) = \mathbb{E}\left[t_{ji}t_{3i}\right] - \mathbb{E}\left[t_{ji}\right]\mathbb{E}\left[t_{3i}\right] = \mathbb{E}\left[t_{3i}\right] - \mathbb{E}\left[t_{ji}\right]\mathbb{E}\left[t_{3i}\right] = P_3\left(1 - P_j\right)$, that is

$$\begin{aligned}
\text{Cov}(t_{ji}, t^*_{3i}) &= \left[\frac{P_3\left(1 - P_j\right) - a_j P_3}{(1 - a_j)(1 - a_k)P_3\left(1 - P_j\right)}\right]\text{Cov}\left(t_{ji}, t_{3i}\right) \\
&= \left[\frac{1 - a_j - P_j}{(1 - a_j)\left(1 - P_j\right)(1 - a_k)}\right]\text{Cov}\left(t_{ji}, t_{3i}\right)
\end{aligned}$$

so that, for $j, k \in \{1, 2\}$ and $j \neq k$,

$$\text{Cov}(t_{ji}, t^*_{3i}) = \zeta_k \eta_j \text{Cov}(t_{ji}, t_{3i})$$

where $\zeta_k = \dfrac{1}{1 - a_k}$ and $\eta_j = \dfrac{1 - a_j - P_j}{1 - a_j}$

The remaining terms in W are determined by covariance between x_i, assumed to be error-free, and underlying (true) program participation status, t^*_{1i} and t^*_{2i}, and the interaction term, t^*_{3i}. Likewise, we can express these terms as functions of misreporting probabilities and sample statistics. For $j \in \{1, 2, 3\}$, it follows that

$$\begin{aligned}
\text{Cov}(x_i, t^*_{ji}) &= \mathbb{E}[x_i t^*_{ji}] - \mathbb{E}[x_i]\mathbb{E}[t^*_{ji}] \\
&= \mathbb{E}[x_i \mathbb{E}[t^*_{ji}|x_i]] - \mathbb{E}[x_i]\mathbb{E}[\exp[t^*_j|x_i]] \quad \text{by the Law of Iterated Expectations} \\
&= \mathbb{E}\left[\frac{x_i \mathbb{E}[t_{ji}|x_i]}{1 - a_j}\right] - \mathbb{E}[x_i]\mathbb{E}\left[\mathbb{E}[\frac{t_{ji}|x_i]}{1 - a_j}\right] \\
&= \frac{\mathbb{E}\left[x_i t_{ji}\right] - \mathbb{E}[x_i]\mathbb{E}\left[t_{ji}\right]}{1 - a_j} \\
&= \frac{\text{Cov}(x_i, t_{ji})}{1 - a_j} = \zeta_j \text{Cov}(x_i, t_{ji})
\end{aligned}$$

We obtain Lemma 2 by combining the results above, giving us the estimator of W, $\hat{W}(a_1, a_2)$, which is a function of misclassification probabilities and other sample statistics that can easily computed in the data.

A.3 Proof of Theorem 1

Proof. *Consistency*: The Adjusted Least Squares Estimator is given by

$$\begin{bmatrix} \hat{\alpha}_{1Adj} \\ \hat{\alpha}_{2Adj} \\ \hat{\alpha}_{3Adj} \\ \hat{\beta}_{Adj} \end{bmatrix} = \begin{bmatrix} \eta_1 \sigma^2_{t_1} & \zeta_2 \sigma_{t_1 t_2} & \eta_1 \zeta_2 \sigma_{t_1 t_3} & \sigma_{t_1 x} \\ \zeta_1 \sigma_{t_2 t_1} & \eta_2 \sigma^2_{t_2} & \eta_2 \zeta_1 \sigma_{t_1 t_3} & \sigma_{t_2 x} \\ \eta_1 \sigma_{t_3 t_1} & \eta_2 \sigma_{t_3 t_2} & \eta_3 \sigma^2_{t_3} & \sigma_{t_3 x} \\ \zeta_1 \sigma_{x t_1} & \zeta_2 \sigma_{x t_2} & \zeta_3 \sigma_{x t_3} & \sigma^2_x \end{bmatrix}^{-1} \begin{bmatrix} \sigma_{y t_1} \\ \sigma_{y t_2} \\ \sigma_{y t_3} \\ \sigma_{yx} \end{bmatrix}$$

which can be expressed in matrix and vector notations as:

$$\hat{\theta}_{Adj} = W\,(a_1, a_2)^{-1}\,\Sigma_{Zy}$$

where $Z = [t_1, t_2, t_3, x]$, $\Sigma_{Zy} = [\sigma_{yt_1}\ \sigma_{yt_2}\ \sigma_{yt_3}\ \sigma_{yx}]'$ and for any covariates, say r_i and s_i, $\sigma_{rs} = \frac{1}{n}\sum_{i=1}^{n}(r_i - \bar{r})(s_i - \bar{s})'$ and $\sigma_r^2 = \frac{1}{n}\sum_{i=1}^{n}(r_i - \bar{r})(r_i - \bar{r})'$where $\bar{r}$ and $\bar{s}$ are the sample mean of r_i and s_i. It follows that $\hat{\theta}_{Adj} = W\,(a_1, a_2)^{-1}\,(\Sigma_{ZZ^*}\theta + \Sigma_{Z\epsilon})$.

We already know from Lemma 2 that Σ_{ZZ^*} can be expressed in terms of misclassification probabilities as $W\,(a_1, a_2)$, so $\hat{\theta}_{Adj} = \theta + W\,(a_1, a_2)^{-1}\,\Sigma_{Z\epsilon}$. Hence, it follows that

$$\hat{\theta}_{Adj} - \theta = W\,(a_1, a_2)^{-1}\,\Sigma_{Z\epsilon}$$

By taking the probability limits and applying Slutsky's Lemma, we have

$$\begin{aligned} \text{plim}\ \hat{\theta}_{Adj} - \theta &= \text{plim}\ W\,(a_1, a_2)^{-1}\ \ \text{plim}\ \Sigma_{Z\epsilon} \\ &= \Gamma^{-1}\Lambda \end{aligned}$$

where, by the Weak Law of Large numbers, we have

$$\Gamma = \begin{bmatrix} \eta_1 \text{Var}(t_{1i}) & \zeta_2 \text{Cov}(t_{1i}, t_{2i}) & \eta_1 \zeta_2 \text{Cov}(t_{1i}, t_{3i}) & \text{Cov}(t_{1i}, x_i) \\ \zeta_1 \text{Cov}(t_{2i}, t_{1i}) & \eta_2 \text{Var}(t_{2i}) & \eta_2 \zeta_1 \text{Cov}(t_{1i}, t_{3i}) & \text{Cov}(t_{2i}, x_i) \\ \eta_1 \text{Cov}(t_{3i}, t_{1i}) & \eta_2 \text{Cov}(t_{3i}, t_{2i}) & \eta_3 \text{Var}(t_{3i}) & \text{Cov}(t_{3i}, x_i) \\ \zeta_1 \text{Cov}(x_i, t_{1i}) & \zeta_2 \text{Cov}(x_i, t_{2i}) & \zeta_3 \text{Cov}(x_i, t_{3i}) & \text{Var}(x_i) \end{bmatrix}$$

and $\Lambda = [\text{Cov}(t_{1i}, \epsilon_i),\ \text{Cov}(t_{2i}, \epsilon_i),\ \text{Cov}(t_{3i}, \epsilon_i),\ \text{Cov}(x_i, \epsilon_i)]'$. By the Law of Iterated Expectations and under Assumption 1,

$\mathbb{E}[\epsilon_i | t_{1i}, t_{2i}, t_{3i}, x_i] = \mathbb{E}[\mathbb{E}[\epsilon_i | t^*_{1i}, t^*_{2i}, t^*_{3i}, x_i] | t_{1i}, t_{2i}, t_{3i}, x_i] = 0$, therefore

$\text{Cov}(t_{1i}, \epsilon_i) = \text{Cov}(t_{2i}, \epsilon_i) = \text{Cov}(t_{3i}, \epsilon_i) = \text{Cov}(x_i, \epsilon_i) = 0$, implying that $\Lambda = 0$, and

$\hat{\theta}_{Adj} - \theta = 0$. This is equivalent to $\theta_{Adj} \xrightarrow{p} \theta$, which translates to, $\alpha_{1Adj} \xrightarrow{p} \alpha_1$, $\alpha_{2Adj} \xrightarrow{p} \alpha_2$, $\alpha_{3Adj} \xrightarrow{p} \alpha_3$, and $\beta_{Adj} \xrightarrow{p} \beta$.

Appendix B. Data Description for Chapter I

B.1 Food Insecurity Score

The food Insecurity score in our analysis represents the USDA's 30-day Adult Food Security Scale obtained from the 10 questions (E2-E9a) included in the last interview of the National Household Food Acquisition and Purchase Survey (FoodAPS) to examine household food security status. The questions take into account severity of conditions and behaviors that characterize food insecurity. We represent the 10 questions in Appendix B Table B1.

Table B1. Food Security Score Questions in National Household Food Acquisition and Purchase Survey (FoodAPS)

Variable	Definition
Question E2	In last 30 days, worried food would run out before we got more money 1 - Often True, 2 - Sometimes True, 3 - Never True
Question E3	Food ran out and had no money to buy more, in last 30 days 1 - Often True, 2 - Sometimes True, 3 - Never True
Question E4	Couldn't afford to eat balanced meals, in last 30 days 1 - Often True, 2 - Sometimes True, 3 - Never True, -997 - Don't Know
Question E5	Adults skipped or cut size of meals b/c not enough money, in last 30 days (Y/N) 0 - No, 1 - Yes, -998- Refused, -996 - Valid Skip
Question E5a	Number of days adults skipped/cut meal size b/c not enough money, last 30 days
Question E6	Eat less than felt you should b/c not enough money, in last 30 days (Y/N) 0 - No, 1 - Yes, -998- Refused, -997 - Don't Know, -996 - Valid Skip
Question E7	Ever hungry but didn't eat b/c not enough money, in last 30 days (Y/N) 0 - No, 1 - Yes, -996 - Valid Skip
Question E8	Lose weight b/c not enough money for food, in last 30 days (Y/N) 0 - No, 1 - Yes, -997 - Don't Know, -996 - Valid Skip
Question E9	Skip food all day b/c not enough money for food, in last 30 days (Y/N) 0 - No, 1 - Yes, -996 - Valid Skip
Question E9a	How often adults skipped food all day b/c not enough money, in last 30 days -997 - Don't Know, -996 - Valid Skip

The Food Security Scale codes "Yes," "Often," "Sometimes," and "three or more days" responses to E2 - E9a questions described above as affirmative responses. The Food Security Score is obtained by summing up the affirmative responses, ranging from 0 to 10. In our analysis, we term it Food Insecurity Score to reflect that 0 represents high food security and increasing values indicate increasing food inadequacy.

B.2 Health Eating Index-2010

National Household Food Purchase and Acquisition Survey (FoodAPS) also aimed to provide data that can be used to evaluate the nutrition quality of food acquired by households. For one week, between April 2012 and January 2013, the survey collected detailed information regarding the types of food obtained by households that can be used in computing the Healthy Eating Index (HEI). The Healthy Eating Index (HEI) measures overall diet quality and the quality of several dietary components, which can be used to examine compliance with the U.S. Dietary Guidelines for Americans (DGAs). Several iterations of HEI have been developed by the U.S. Department of Health and Human Services National Cancer Institute (NCI) and the U.S. Department of Agriculture (USDA) researchers since 2005. Following Mancino et al. (2018), we consider 2010 Health Eating Index (HEI-2010) scores to assess the quality of food items reported in FoodAPS. HEI-2010 scores range from 0 to 100 and have 12 components, as presented in Appendix Table B2 . The calculation of HEI-2010 accounts for the variation in individual total caloric needs. The process matches the reported food items in FoodAPS with Food Pattern Equivalent Database for each food item and the USDA nutrient food code from USDA Food and Nutrient Database for Dietary Studies (FNDDS). We obtain the method, and the code to compute the 2010 HEI from Mancino et al. (2018).

Table B2. HEI–2010 Components and Scoring Standards

Component	Maximum Points	Standard For Maximum Score	Standard For Minimum Score of Zero
Adequacy:			
Total Fruit[a]	5	≥ 0.8 cup equiv. per 1,000 kcal	No Fruit
Whole Fruit[b]	5	≥ 0.4 cup equiv. per 1,000 kcal	No Whole Fruit
Total Vegetables[c]	5	≥ 1.1 cup equiv. per 1,000 kcal	No Vegetables
Greens and Beans[c]	5	≥ 0.2 cup equiv. per 1,000 kcal	No Dark Green Vegetables or Beans and Peas
Whole Grains	10	≥ 1.5 oz equiv. per 1,000 kcal	No Whole Grains
Dairy[d]	10	≥ 1.3cup equiv. per 1,000 kcal	No Dairy
Total Protein Foods	5	≥ 2.5 oz equiv. per 1,000 kcal	No Protein Foods
Seafood and Plant Proteins[e, f]	5	≥ 0.8 oz equiv. per 1,000 kcal	No Seafood or Plant Proteins
Fatty Acids[g]	10	(PUFAs + MUFAs)/SFAs ≥ 2.5	(PUFAs + MUFAs)/SFAs ≤ 1.2
Moderation:			
Refined Grains	10	≤ 1.8 oz equiv. per 1,000 kcal	≥ 4.3 oz equiv. per 1,000 kcal
Sodium	10	≤ 1.1 gram per 1,000 kcal	≥ 2.0 grams per 1,000 kcal
Empty Calories[h]	20	≤ 19% of energy	≥ 50% of energy

Note: Intakes between the minimum and maximum standards are scored proportionately. The total HEI score is the sum of the adequacy components (i.e. foods to eat more of for good health) and moderation components (i.e. foods to limit for good health).

[a] Includes 100% fruit juice.

[b] Includes all forms except juice.

[c] Includes any beans and peas not counted as Total Protein Foods.

[d] Includes all milk products, such as fluid milk, yogurt, and cheese, and fortified soy beverages.

[e] Beans and peas are included here (and not with vegetables) when the Total Protein Foods standard is otherwise not met.

[f] Includes seafood, nuts, seeds, soy products (other than beverages) as well as beans and peas counted as Total Protein Foods.

[g] Ratio of poly- and monounsaturated fatty acids (PUFAs and MUFAs) to saturated fatty acids (SFAs).

[h] Calories from solid fats, alcohol, and added sugars; threshold for counting alcohol is > 13 grams/1000 kcal.

Appendix C. Additional Tables and Figures for Chapter I

Table C1. Impact of Supplemental Nutrition Assistance Program (SNAP) and Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) on Healthy Eating - Adequacy Component Using Misclassification Probabilities from Validation Data

	OLS	Adj.LS	OLS	Adj.LS	OLS	Adj.LS
Panel 1						
Dependent Variables:	Total Fruit		Whole Fruit		Total Vegetables	
SNAP	0.065	-0.022	-0.331	-0.038	-0.013	0.048
	(0.273)	(0.047)	(0.317)	(0.053)	(0.255)	(0.043)
WIC	0.089	-0.043	-0.475	-0.003	0.084	0.044
	(0.289)	(0.058)	(0.336)	(0.063)	(0.270)	(0.055)
SNAP × WIC	0.044	0.003	0.470	0.048	-0.212	-0.020
	(0.342)	(0.032)	(0.398)	(0.038)	(0.320)	(0.033)
Panel 2						
Dependent Variables:	Greens and Beans		Whole Grains		Dairy	
SNAP	-0.543*	0.008	-0.263	-0.042	-0.184	-0.155
	(0.296)	(0.051)	(0.395)	(0.068)	(0.551)	(0.099)
WIC	-0.114	-0.104*	-0.133	-0.096	-0.486	-0.138
	(0.314)	(0.060)	(0.419)	(0.080)	(0.584)	(0.127)
SNAP × WIC	0.431	0.044	0.420	0.041	0.886	0.085
	(0.371)	(0.036)	(0.496)	(0.048)	(0.691)	(0.070)
Panel 3						
Dependent Variables:	Total Protein Foods		Seafood and Plant Proteins		Fatty Acids	
SNAP	-0.129	-0.049	0.112	-0.016	-0.930	0.105
	(0.247)	(0.040)	(0.296)	(0.053)	(0.579)	(0.107)
WIC	-0.254	-0.032	0.146	-0.038	-0.160	-0.050
	(0.262)	(0.057)	(0.314)	(0.063)	(0.614)	(0.129)
SNAP × WIC	0.342	0.033	-0.027	-0.004	0.317	0.036
	(0.310)	(0.033)	(0.371)	(0.037)	(0.727)	(0.073)

Note: Standard errors in parenthesis and bootstrapped. The analytical sample comes from FoodAPS and consists of households with a pregnant woman or child under five years old and below 130% of the federal poverty threshold. The probability of false negatives in SNAP comes from Courtemanche et al. (2019), $a_1 = 32.31\%$, and in WIC comes from Fox and Hokayem (2022), $a_2 = 41.5\%$. Regressors not reported include primary respondent education, ethnicity, race, employment, marital status, age, rural household residency, indicator and count of children below the age of five, number of children, income to poverty ratio, primary store distance, whether the primary store is SNAP authorized, whether the household has any vehicle and whether the household is renting the house. Significance codes: *** p<0.01, ** p<0.05, * p<0.1

Table C2. Impact of Supplemental Nutrition Assistance Program (SNAP) and Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) on Healthy Eating - Moderation Component Using Misclassification Probabilities from Validation Data

	OLS	Adj.LS	OLS	Adj.LS	OLS	Adj.LS
Dependent Variables:	Refined Grains		Sodium		Empty Calories	
SNAP	0.422	-0.120	0.199	0.076	-1.506	-0.147
	(0.621)	(0.115)	(0.627)	(0.108)	(0.987)	(0.181)
WIC	-0.308	0.029	0.311	0.075	-0.667	-0.439**
	(0.658)	(0.134)	(0.664)	(0.136)	(1.046)	(0.222)
SNAP × WIC	0.202	0.017	-0.529	-0.051	1.985	0.195
	(0.779)	(0.074)	(0.787)	(0.072)	(1.239)	(0.127)

Note: Standard errors in parenthesis and bootstrapped. The analytical sample comes from FoodAPS and consists of households with a pregnant woman or child under five years old and below 130% of the federal poverty threshold. The probability of false negatives in SNAP comes from Courtemanche et al. (2019), $a_1 = 32.31\%$, and in WIC comes from Fox and Hokayem (2022), $a_2 = 41.5\%$. Regressors not reported include primary respondent education, ethnicity, race, employment, marital status, age, rural household residency, indicator and count of children below the age of five, number of children, income to poverty ratio, primary store distance, whether the primary store is SNAP authorized, whether the household has any vehicle and whether the household is renting the house. Significance codes: *** p<0.01, ** p<0.05, * p<0.1.

Table C3. Impact of Supplemental Nutrition Assistance Program (SNAP) and Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) on Healthy Eating - Adequacy Component Using Estimated Misclassification Probabilities

	OLS	Adj.LS	OLS	Adj.LS	OLS	Adj.LS
Panel 1						
Dependent Variables:	Total Fruit		Whole Fruit		Total Vegetables	
SNAP	0.065	-0.025	-0.331	-0.042	-0.013	0.055
	(0.273)	(0.053)	(0.317)	(0.060)	(0.255)	(0.049)
WIC	0.089	-0.047	-0.475	-0.005	0.084	0.048
	(0.289)	(0.063)	(0.336)	(0.069)	(0.270)	(0.060)
SNAP × WIC	0.044	0.004	0.470	0.055	-0.212	-0.023
	(0.342)	(0.037)	(0.398)	(0.043)	(0.320)	(0.037)
Panel 2						
Dependent Variables:	Greens and Beans		Whole Grains		Dairy	
SNAP	-0.543*	0.011	-0.263	-0.047	-0.184	-0.176
	(0.296)	(0.059)	(0.395)	(0.077)	(0.551)	(0.113)
WIC	-0.114	-0.114*	-0.133	-0.105	-0.486	-0.151
	(0.314)	(0.066)	(0.419)	(0.088)	(0.584)	(0.139)
SNAP × WIC	0.431	0.050	0.420	0.047	0.886	0.097
	(0.371)	(0.042)	(0.496)	(0.056)	(0.691)	(0.081)
Panel 3						
Dependent Variables:	Total Protein Foods		Seafood and Plant Proteins		Fatty Acids	
SNAP	-0.129	-0.055	0.112	-0.018	-0.930	0.123
	(0.247)	(0.046)	(0.296)	(0.060)	(0.579)	(0.122)
WIC	-0.254	-0.035	0.146	-0.041	-0.160	-0.054
	(0.262)	(0.062)	(0.314)	(0.069)	(0.614)	(0.140)
SNAP × WIC	0.342	0.038	-0.027	-0.005	0.317	0.042
	(0.310)	(0.038)	(0.371)	(0.043)	(0.727)	(0.084)

Note: Standard errors in parenthesis and bootstrapped. The analytical sample comes from FoodAPS and consists of households with a pregnant woman or child under five years old and below 130% of the federal poverty threshold. The probability of false negatives in SNAP is $\hat{a}_1 = 29.7\%$, and in WIC is $\hat{a}_2 = 40.1\%$. The misclassification probabilities are estimated by extending parametric procedure of Hausman et al. (1998) to bivariate models. Regressors not reported include primary respondent education, ethnicity, race, employment, marital status, age, rural household residency, indicator and count of children below the age of five, number of children, income to poverty ratio, primary store distance, whether the primary store is SNAP authorized, whether the household has any vehicle and whether the household is renting the house. Significance codes: *** p<0.01, ** p<0.05, * p<0.1

Table C4. Impact of Supplemental Nutrition Assistance Program (SNAP) and Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) on Healthy Eating - Moderation Component Using Estimated Misclassification Probabilities

	OLS	Adj.LS	OLS	Adj.LS	OLS	Adj.LS
Dependent Variables:	Refined Grains		Sodium		Empty Calories	
SNAP	0.422	-0.139	0.199	0.086	-1.506	-0.164
	(0.621)	(0.131)	(0.627)	(0.123)	(0.987)	(0.206)
WIC	-0.308	0.031	0.311	0.083	-0.667	-0.479**
	(0.658)	(0.146)	(0.664)	(0.148)	(1.046)	(0.242)
SNAP × WIC	0.202	0.020	-0.529	-0.059	1.985	0.225
	(0.779)	(0.085)	(0.787)	(0.083)	(1.239)	(0.146)

Note: Standard errors in parenthesis and bootstrapped. The analytical sample comes from FoodAPS and consists of households with a pregnant woman or child under five years old and below 130% of the federal poverty threshold. The probability of false negatives in SNAP is $\hat{a}_1 = 29.7\%$, and in WIC is $\hat{a}_2 = 40.1\%$. The misclassification probabilities are estimated by extending parametric procedure of Hausman et al. (1998) to bivariate models. Regressors not reported include primary respondent education, ethnicity, race, employment, marital status, age, rural household residency, indicator and count of children below the age of five, number of children, income to poverty ratio, primary store distance, whether the primary store is SNAP authorized, whether the household has any vehicle and whether the household is renting the house. Significance codes: *** p<0.01, ** p<0.05, * p<0.1

Appendix D. Mathematical Proofs for Chapter II

D.1 Proof of Theorem 2

Proof. OLS estimation biasdness: least square estimate β_{LS} in the regression equation of interest model (12) is given by equation (18). That is,

$$\widehat{\beta}_{LS} = (\tilde{T}'\tilde{M}\tilde{T})^{-1}\tilde{T}'\tilde{M}\tilde{y}$$

where $\tilde{M} = I - \tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'$ and for any variable H, $\tilde{H} = RH$. By Cauchy-Shwarz inequality, It follows that

$$\widehat{\beta}_{LS} = (\tilde{T}'\tilde{M}\tilde{T})^{-1}\tilde{T}'\tilde{M}\left(\tilde{X}\gamma + \tilde{T}\beta + \tilde{\boldsymbol{\varepsilon}}\right)$$

This implies that $\widehat{\beta}_{LS} = \beta + (\tilde{T}'\tilde{M}\tilde{T})^{-1}\tilde{T}'\tilde{M}\tilde{\boldsymbol{\varepsilon}}$, and OLS bias is given by $\widehat{\beta}_{LS} - \beta = (\tilde{T}'\tilde{M}\tilde{T})^{-1}\tilde{T}'\tilde{M}\tilde{\boldsymbol{\varepsilon}}$. By Cauchy-Schwarz inequality, $\tilde{T}'\tilde{M}\tilde{T}$ is positive-semi definite, hence the bias term in complete-case analysis is driven by $\tilde{T}'\tilde{M}\tilde{\boldsymbol{\varepsilon}}$.

Inconsistency: $\widehat{\beta}_{LS} = \beta + (\tilde{T}'\tilde{M}\tilde{T})^{-1}\tilde{T}'\tilde{M}\tilde{\boldsymbol{\varepsilon}}$ can be expressed as

$$\widehat{\beta}_{LS} = \beta + \left(\frac{\tilde{T}'\tilde{M}\tilde{T}}{n}\right)^{-1}\frac{\tilde{T}'\tilde{M}\tilde{\boldsymbol{\varepsilon}}}{n}$$

By taking the probability limit and applying the Slutsky's lemma, we have

$$\text{plim}\,\widehat{\beta}_{LS} = \beta + \text{plim}\left(\frac{\tilde{T}'\tilde{M}\tilde{T}}{n}\right)^{-1}\text{plim}\,\frac{\tilde{T}'\tilde{M}\tilde{\boldsymbol{\varepsilon}}}{n} \qquad \text{(D1)}$$

Using the definition $\tilde{M}$, we can decompose $\text{plim}\frac{\tilde{T}'\tilde{M}\tilde{T}}{n}$ to

$$\begin{aligned}\text{plim}\,\frac{\tilde{T}'\tilde{M}\tilde{T}}{n} &= \text{plim}\,\frac{\tilde{T}'\left(\tilde{I} - \tilde{X}\left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\right)\tilde{T}}{n} \\ &= \text{plim}\,\frac{\tilde{T}'\tilde{T}}{n} - \text{plim}\frac{\tilde{T}'\tilde{X}\left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\tilde{T}}{n}\end{aligned}$$

By the Weak Law of Large Numbers, we have

$$\frac{\tilde{T}'\tilde{M}\tilde{T}}{n} \xrightarrow{p} \mathbb{E}\left(T_i R_i\right) - \mathbb{E}\left(T_i R_i \mathrm{x}_i'\right) \mathbb{E}\left(\mathrm{x}_i \mathrm{x}_i' R_i\right)^{-1} \mathbb{E}\left(T_i R_i \mathrm{x}_i\right)$$

and Continuous Mapping Theorem implies that

$$\left(\frac{\tilde{T}'\tilde{M}\tilde{T}}{n}\right)^{-1} \xrightarrow{p} \left(\mathbb{E}\left(T_i R_i\right) - \mathbb{E}\left(T_i R_i \mathrm{x}_i'\right) \mathbb{E}\left(\mathrm{x}_i \mathrm{x}_i' R_i\right)^{-1} \mathbb{E}\left(T_i R_i \mathrm{x}_i\right)\right)^{-1} \tag{D2}$$

Where $\mathbb{E}\left(T_i R_i\right) - \mathbb{E}\left(T_i R_i \mathrm{x}_i'\right) \mathbb{E}\left(\mathrm{x}_i \mathrm{x}_i' R_i\right)^{-1} \mathbb{E}\left(T_i R_i \mathrm{x}_i\right) > 0$ by Cauchy-Schwarz inequality. Likewise, we can decompose $\operatorname{plim}\frac{\tilde{T}'\tilde{M}\tilde{\boldsymbol{\varepsilon}}}{n}$ into

$$\begin{aligned}
\operatorname{plim}\frac{\tilde{T}'\tilde{M}\tilde{\boldsymbol{\varepsilon}}}{n} &= \operatorname{plim}\frac{\tilde{T}'\left(\tilde{I} - \tilde{X}\left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\right)\tilde{\boldsymbol{\varepsilon}}}{n} \\
&= \operatorname{plim}\frac{\tilde{T}'\tilde{\boldsymbol{\varepsilon}}}{n} - \operatorname{plim}\frac{\tilde{T}'\tilde{X}\left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\tilde{\boldsymbol{\varepsilon}}}{n}
\end{aligned}$$

Following similar arguments as above, we have

$$\frac{\tilde{T}'\tilde{M}\tilde{\boldsymbol{\varepsilon}}}{n} \xrightarrow{p} \mathbb{E}\left(T_i \varepsilon_i R_i\right) - \mathbb{E}\left(T_i R_i \mathrm{x}_i'\right) \mathbb{E}\left(\mathrm{x}_i \mathrm{x}_i' R_i\right)^{-1} \mathbb{E}\left(\mathrm{x}_i R_i \varepsilon_i\right)$$

Taking the participation regression model given by equation (13) and the missing mechanism described by equation (14), we have:

$$\begin{aligned}
\mathbb{E}\left[T_i \varepsilon_i R_i\right] &= \mathbb{E}\left[\varepsilon_i \mathbb{I}\left(\mathrm{z_i}\delta + u_i \geq 0, \mathrm{w_i}\theta + v_i \geq 0\right)\right] \\
&= \mathbb{E}\left[\Pr\left(u_i \geq -\mathrm{z_i}\delta, v_i \geq -\mathrm{w_i}\theta, \rho\right) \mathbb{E}\left(\varepsilon_i | u_i \geq -\mathrm{z_i}\delta, v_i \geq -\mathrm{w_i}\theta\right)\right]
\end{aligned}$$

Let $\phi(\cdot)$ and $\Phi(\cdot)$ be PDF and CDF of standard normal distribution. By bivariate normality of (u_i, v_i), $\Pr\left(u_i \geq -\mathrm{z_i}\delta, v_i \geq -\mathrm{w_i}\theta, \rho\right) = \Phi\left(\mathrm{z_i}\delta, \mathrm{w_i}\theta, \rho\right)$. We can obtain the conditional mean of the error term (ε_i) by integrating the joint distribution of $(\varepsilon_i, u_i, v_i)$ which is trivariate normal. This gives us:

$$\mathbb{E}\left[\varepsilon_i | u_i \geq -\mathrm{z_i}\delta, v_i \geq -\mathrm{w_i}\theta\right] = \sigma\rho_u \frac{\phi\left(\mathrm{z}_i'\delta\right)\Phi\left(\frac{\mathrm{w}_i'\theta - \rho \mathrm{z}_i'\delta}{\sqrt{1-\rho^2}}\right)}{\Phi\left(\mathrm{z_i}\delta, \mathrm{w_i}\theta, \rho\right)} + \sigma\rho_v \frac{\phi\left(\mathrm{w}_i'\theta\right)\Phi\left(\frac{\mathrm{z}_i'\delta - \rho \mathrm{w}_i'\theta}{\sqrt{1-\rho^2}}\right)}{\Phi\left(\mathrm{z_i}\delta, \mathrm{w_i}\theta, \rho\right)}$$

This implies,

$$\mathbb{E}\left[T_i\varepsilon_i R_i\right] = \mathbb{E}\left[\sigma\rho_u\phi\left(\mathrm{z}_i'\delta\right)\Phi\left(\frac{\mathrm{w}_i'\theta - \rho \mathrm{z}_i'\delta}{\sqrt{1-\rho^2}}\right) + \sigma\rho_v\phi\left(\mathrm{w}_i'\theta\right)\Phi\left(\frac{\mathrm{z}_i'\delta - \rho \mathrm{w}_i'\theta}{\sqrt{1-\rho^2}}\right)\right]$$

Likewise, we can express $\mathbb{E}\left(\mathrm{x}_i R_i \varepsilon_i\right)$ as

$$\begin{aligned}
\mathbb{E}\left[\mathrm{x}_i R_i \varepsilon_i\right] &= \mathbb{E}\left[\mathrm{x}_i \mathbb{I}\left(\mathrm{w}_i'\theta + v_i \geq 0\right)\varepsilon_i\right] \\
&= \mathbb{E}\left[\mathrm{x}_i \Pr\left(v_i \geq -\mathrm{w}_i'\theta\right)\mathbb{E}\left[\varepsilon_i | v_i \geq -\mathrm{w}_i'\theta\right]\right] \\
&= \mathbb{E}\left[\mathrm{x}_i \Phi\left(\mathrm{w}_i'\theta\right)\mathbb{E}\left[\varepsilon_i | v_i \geq -\mathrm{w}_i'\theta\right]\right] \\
&= \mathbb{E}\left[\mathrm{x}_i \Phi\left(\mathrm{w}_i'\theta\right)\sigma\rho_v \frac{\phi\left(\mathrm{w}_i'\theta\right)}{\Phi\left(\mathrm{w}_i'\theta\right)}\right] \\
&= \sigma\rho_v \mathbb{E}\left[\mathrm{x}_i \phi\left(\mathrm{w}_i'\theta\right)\right]
\end{aligned}$$

Hence,

$$\begin{aligned}
\frac{\tilde{T}'\tilde{M}\tilde{\boldsymbol{\varepsilon}}}{n} \xrightarrow{p} \sigma\rho_u \mathbb{E}\left[\phi\left(\mathrm{z}_i'\delta\right)\Phi\left(\frac{\mathrm{w}_i'\theta - \rho \mathrm{z}_i'\delta}{\sqrt{1-\rho^2}}\right)\right] + \sigma\rho_v \mathbb{E}\left[\phi\left(\mathrm{w}_i'\theta\right)\Phi\left(\frac{\mathrm{z}_i'\delta - \rho \mathrm{w}_i'\theta}{\sqrt{1-\rho^2}}\right)\right] \\
- \sigma\rho_v \mathbb{E}\left(T_i R_i \mathrm{x}_i'\right)\mathbb{E}\left(\mathrm{x}_i \mathrm{x}_i' R_i\right)^{-1}\mathbb{E}\left[\mathrm{x}_i \phi\left(\mathrm{w}_i'\theta\right)\right]
\end{aligned} \tag{D3}$$

Combining Equations (D1), (D2), and (D3), we obtain the results in Theorem 2

$$\beta_{LS} \xrightarrow{p} \beta + \sigma\left[\varphi_u A(\rho) + \varphi_v B(\rho)\right]$$

where

$$A(\rho) = \frac{\mathbb{E}\left[\phi\left(\mathrm{z}_i'\delta\right)\Phi\left(\frac{\mathrm{w}_i'\theta - \rho \mathrm{z}_i'\delta}{\sqrt{1-\rho^2}}\right)\right]}{\mathbb{E}(T_i R_i) - \mathbb{E}(T_i R_i \mathrm{x}_i)\mathbb{E}(\mathrm{x}_i \mathrm{x}_i' R_i)^{-1}\mathbb{E}[T_i R_i \mathrm{x}_i']},$$

$$B(\rho) = \frac{\mathbb{E}\left[\phi\left(\mathrm{w}_i'\theta\right)\Phi\left(\frac{\mathrm{z}_i'\delta - \rho \mathrm{w}_i'\theta}{\sqrt{1-\rho^2}}\right)\right] - \mathbb{E}(T_i R_i \mathrm{x}_i')\mathbb{E}(\mathrm{x}_i \mathrm{x}_i' R_i)^{-1}\mathbb{E}[\mathrm{x}_i \phi\left(\mathrm{w}_i'\theta\right)]}{\mathbb{E}(T_i R_i) - \mathbb{E}(T_i R_i \mathrm{x}_i)\mathbb{E}(\mathrm{x}_i \mathrm{x}_i' R_i)^{-1}\mathbb{E}[T_i R_i \mathrm{x}_i']}$$

Proof. Instrument variable estimation inconsistency: The instrumental variable estimator of β as shown in equation (19) is given by

$$\widehat{\beta}_{IV} = \left(\tilde{\Phi}\tilde{M}\tilde{\Phi}\right)^{-1}\tilde{\Phi}\tilde{M}\tilde{y}$$

Where $\tilde{\Phi}_i = \Phi(\mathrm{z}_i \hat{\boldsymbol{\delta}}_c)$ are predicted probabilities of participation obtained from the first stage estimation, $\Pr[\tilde{T}_i = 1|\mathrm{z}_i, R_i] = \Phi(\mathrm{z}_i\delta)$. It follows that,

$$\begin{aligned}
\widehat{\beta}_{IV} &= \left(\tilde{\Phi}\tilde{M}\tilde{\Phi}\right)^{-1}\tilde{\Phi}\tilde{M}\tilde{y} \\
&= \left(\tilde{\Phi}\tilde{M}\tilde{\Phi}\right)^{-1}\tilde{\Phi}\tilde{M}\left(\tilde{X}\gamma + \tilde{T}\beta + \tilde{\boldsymbol{\varepsilon}}\right) \\
&= \beta + \left(\tilde{\Phi}\tilde{M}\tilde{\Phi}\right)^{-1}\tilde{\Phi}\tilde{M}\tilde{\boldsymbol{\varepsilon}}
\end{aligned}$$

which can be expressed as

$$\widehat{\beta}_{IV} = \beta + \left(\frac{\tilde{\Phi}\tilde{M}\tilde{\Phi}}{n}\right)^{-1}\frac{\tilde{\Phi}\tilde{M}\tilde{\boldsymbol{\varepsilon}}}{n}$$

Takin probability limit and using Slutsky Lemma, we have

$$\text{plim}\widehat{\beta}_{IV} = \beta + \text{plim}\left(\frac{\tilde{\Phi}\tilde{M}\tilde{\Phi}}{n}\right)^{-1}\text{plim}\frac{\tilde{\Phi}\tilde{M}\tilde{\boldsymbol{\varepsilon}}}{n} \qquad \text{(D4)}$$

Expanding projection matrix $\tilde{M}$, we can decompose $\text{plim}\frac{\tilde{\Phi}\tilde{M}\tilde{\Phi}}{n}$ as follows

$$\begin{aligned}\text{plim}\frac{\tilde{\Phi}\tilde{M}\tilde{\Phi}}{n} &= \text{plim}\,\frac{\tilde{\Phi}'\left(\tilde{I}-\tilde{X}\left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\right)\tilde{\Phi}}{n} \\ &= \text{plim}\,\frac{\tilde{\Phi}'\tilde{\Phi}}{n} - \text{plim}\frac{\tilde{\Phi}'\tilde{X}\left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\tilde{\Phi}}{n}\end{aligned}$$

By applying Weak Law of Large Number and expressing the first stage estimates in full, it follows that

$$\begin{aligned}\text{plim}\frac{\tilde{\Phi}\tilde{M}\tilde{\Phi}}{n} &= \mathbb{E}\left[\Phi(\mathrm{z}_i'\delta_c)^2R_i\right] - \mathbb{E}\left[\Phi(\mathrm{z}_i'\delta_c)R_i\mathrm{x}_i'\right]\mathbb{E}\left[\mathrm{x}_i\mathrm{x}_i'R_i\right]^{-1}\mathbb{E}\left[\Phi(\mathrm{z}_i'\delta_c)R_i\mathrm{x}_i\right] \\ &= \mathbb{E}\left[\Phi(\mathrm{z}_i'\delta_c)^2\Phi\left(\mathrm{w}_i'\theta\right)\right] \\ &\quad - \mathbb{E}\left[\Phi(\mathrm{z}_i'\delta_c)\Phi\left(\mathrm{w}_i'\theta\right)\mathrm{x}_i'\right]\mathbb{E}\left[\mathrm{x}_i\mathrm{x}_i'\Phi\left(\mathrm{w}_i'\theta\right)\right]^{-1}\mathbb{E}\left[\Phi(\mathrm{z}_i'\delta_c)\Phi\left(\mathrm{w}_i'\theta\right)\mathrm{x}_i\right]\end{aligned}$$

Using Continuous mapping theorem, we have

$$\begin{aligned}\left(\frac{\tilde{\Phi}\tilde{M}\tilde{\Phi}}{n}\right)^{-1} \xrightarrow{p} \Big[&\mathbb{E}\left[\Phi(\mathrm{z}_i'\delta_c)^2\Phi\left(\mathrm{w}_i'\theta\right)\right] \\ &- \mathbb{E}\left[\Phi(\mathrm{z}_i'\delta_c)\Phi\left(\mathrm{w}_i'\theta\right)\mathrm{x}_i'\right]\mathbb{E}\left[\mathrm{x}_i\mathrm{x}_i'\Phi\left(\mathrm{w}_i'\theta\right)\right]^{-1}\mathbb{E}\left[\Phi(\mathrm{z}_i'\delta_c)\Phi\left(\mathrm{w}_i'\theta\right)\mathrm{x}_i\right]\Big]^{-1}\end{aligned} \tag{D5}$$

Similarly, we can express $\text{plim}\frac{\tilde{\Phi}\tilde{M}\tilde{\varepsilon}}{n}$ as

$$\begin{aligned}\text{plim}\frac{\tilde{\Phi}\tilde{M}\tilde{\varepsilon}}{n} &= \text{plim}\,\frac{\tilde{\Phi}'\left(\tilde{I}-\tilde{X}\left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\right)\tilde{\varepsilon}}{n} \\ &= \text{plim}\,\frac{\tilde{\Phi}'\tilde{\varepsilon}}{n} - \text{plim}\frac{\tilde{\Phi}'\tilde{X}\left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\tilde{\varepsilon}}{n}\end{aligned}$$

Following similar argument as above, we have

$$\text{plim}\frac{\tilde{\Phi}\tilde{M}\tilde{\varepsilon}}{n} = \mathbb{E}\left[\Phi(\mathrm{z}_i'\delta_c)\varepsilon_iR_i\right] - \mathbb{E}\left[\Phi(\mathrm{z}_i'\delta_c)R_i\mathrm{x}_i'\right]\mathbb{E}\left[\mathrm{x}_i\mathrm{x}_i'R_i\right]^{-1}\mathbb{E}\left[\mathrm{x}_iR_i\varepsilon_i\right]$$

Where

$$\begin{aligned}\mathbb{E}\left[\Phi(z_i'\delta_c)R_i\varepsilon_i\right] &= \mathbb{E}\left[\Phi(z_i'\delta_c)\Phi(w_i'\theta)\mathbb{E}[\varepsilon_i|v_i \geq -w_i'\theta]\right] \\ &= \mathbb{E}\left[\Phi(z_i'\delta_c)\Phi(w_i'\theta)\sigma\rho_v\frac{\phi\left(w_i\theta\right)}{\Phi(w_i'\theta)}\right] \\ &= \sigma\rho_v\mathbb{E}\left[\Phi(z_i'\delta_c)\phi\left(w_i\theta\right)\right]\end{aligned}$$

And

$$\begin{aligned}\mathbb{E}\left[x_iR_i\varepsilon_i\right] &= \mathbb{E}\left[x_i\Phi(w_i'\theta)\mathbb{E}[\varepsilon_i|v_i \geq -w_i'\theta]\right] \\ &= \mathbb{E}\left[x_i\Phi(w_i'\theta)\sigma\rho_v\frac{\phi\left(w_i\theta\right)}{\Phi(w_i'\theta)}\right] \\ &= \sigma\rho_v\mathbb{E}\left[x_i\phi\left(w_i\theta\right)\right]\end{aligned}$$

Now we have

$$\begin{aligned}\frac{\tilde{\Phi}\tilde{M}\tilde{\varepsilon}}{n}\xrightarrow{p}\sigma\rho_v\Big[&\mathbb{E}\left[\Phi(z_i'\delta_c)\phi\left(w_i\theta\right)\right] \\ &-\mathbb{E}\left[\Phi(z_i'\delta_c)\Phi\left(w_i\theta\right)x_i'\right]\mathbb{E}\left[x_ix_i'\Phi\left(w_i\theta\right)\right]^{-1}\mathbb{E}\left[x_i\phi\left(w_i\theta\right)\right]\Big]\end{aligned} \tag{D6}$$

The results in Theorem 2 are obtained by combining Equations (D4), (D5, and (D6), that is

$$\beta_{IV}\xrightarrow{p}\beta+\sigma\varphi_vC \tag{D7}$$

where C is given by:

$$C=\frac{\mathbb{E}\left[\Phi(z_i'\delta_c)\phi\left(w_i'\theta\right)\right]-\mathbb{E}\left[\Phi(z_i'\delta_c)\Phi\left(w_i'\theta\right)x_i'\right]\mathbb{E}\left[x_ix_i'\Phi(w_i\theta)\right]^{-1}\mathbb{E}\left[x_i\phi\left(w_i'\theta\right)\right]}{\mathbb{E}\left[\Phi(z_i'\delta_c)^2\Phi\left(w_i'\theta\right)\right]-\mathbb{E}\left[\Phi(z_i'\delta_c)\Phi\left(w_i'\theta\right)x_i'\right]\mathbb{E}\left[x_ix_i'\Phi(w_i\theta)\right]^{-1}\mathbb{E}\left[\Phi(z_i'\delta_c)\Phi\left(w_i'\theta\right)x_i\right]}$$

D.2 Proof of Theorem 3

Proof. Using predicted probabilities from the second step, $\hat{T}_i$, the three step estimator of β can be expressed as:

$$\begin{aligned}\beta_{3s} &= (\hat{T}'M\hat{T})^{-1}\hat{T}'My \\ &= (\hat{T}'M\hat{T})^{-1}\hat{T}'MT\beta + (\hat{T}'M\hat{T})^{-1}\hat{T}'M\boldsymbol{\epsilon}\end{aligned}$$

It follows that

$$\beta_{3s} = \left(\frac{\hat{T}'M\hat{T}}{n}\right)^{-1}\frac{\hat{T}'MT}{n}\beta + \left(\frac{\hat{T}'M\hat{T}}{n}\right)^{-1}\frac{\hat{T}'M\boldsymbol{\epsilon}}{n} \tag{D8}$$

Consider the first term on the RHS of equation (D8). By finite variance assumption, weak law of large numbers, continuity of $\Phi(\cdot)$, and consistency of $\hat{\delta}$ in the second step:

$$\text{plim}\frac{\hat{T}'M\hat{\boldsymbol{T}}}{n} \xrightarrow{p} \mathbb{E}\left[\Phi(\mathrm{z}_i'\delta)^2\right] - \mathbb{E}\left[\Phi(\mathrm{z}_i'\delta)\mathrm{x}_i\right]\mathbb{E}[\mathrm{x}_i'\mathrm{x}_i]^{-1}\mathbb{E}\left[\mathrm{x}_i\Phi(\mathrm{z}_i'\delta)\right]$$

and

$$\begin{aligned}\text{plim}\frac{\hat{T}'M\boldsymbol{T}}{n} &\xrightarrow{p} \mathbb{E}\left[\Phi(\mathrm{z}_i'\delta)T_i\right] - \mathbb{E}\left[\Phi(\mathrm{z}_i'\delta)\mathrm{x}_i\right]\mathbb{E}[\mathrm{x}_i'\mathrm{x}_i]^{-1}\mathbb{E}\left[\mathrm{x}_iT_i\right] \\ &= \mathbb{E}\left[\Phi(\mathrm{z}_i'\delta)\mathbb{E}[T_i|\mathrm{z}_i]\right] - \mathbb{E}\left[\Phi(\mathrm{z}_i'\delta)\mathrm{x}_i\right]\mathbb{E}[\mathrm{x}_i'\mathrm{x}_i]^{-1}\mathbb{E}\left[\mathrm{x}_i\mathbb{E}[T_i|\mathrm{z}_i]\right] \\ &= \mathbb{E}\left[\Phi(\mathrm{z}_i'\delta)^2\right] - \mathbb{E}\left[\Phi(\mathrm{z}_i'\delta)\mathrm{x}_i\right]\mathbb{E}[\mathrm{x}_i'\mathrm{x}_i]^{-1}\mathbb{E}\left[\mathrm{x}_i\Phi(\mathrm{z}_i'\delta)\right]\end{aligned}$$

Note that participation status is determined by a set of instruments, z_i, as described in equation (13) and $\mathbb{E}[T_i|\mathrm{z}_i] = \Phi(\mathrm{z}_i'\delta)$. It follows that:

$$\left(\frac{\hat{T}'M\hat{T}}{n}\right)^{-1}\frac{\hat{T}'MT}{n} \xrightarrow{p} 1$$

Following steps and assumptions above (continuity of $\Phi(\cdot)$, consistency of $\hat{\delta}$, and the weak law of large numbers):

$$\text{plim}\frac{\hat{T}'M\boldsymbol{\epsilon}}{n}\xrightarrow{p}\mathbb{E}\left[\Phi(\mathrm{z}_i'\delta)\epsilon_i\right]-\mathbb{E}\left[\Phi(\mathrm{z}_i'\delta)\mathrm{x}_i\right]\mathbb{E}[\mathrm{x}_i'\mathrm{x}_i]^{-1}\mathbb{E}\left[\mathrm{x}_i\epsilon_i\right]$$

By exogeneity of x and z:

$$\mathbb{E}\left[\Phi(\mathrm{z}_i'\delta)\epsilon_i\right]=\mathbb{E}\left[\Phi(\mathrm{z}_i'\delta)\mathbb{E}[\epsilon_i|\mathrm{z}_i]\right]=0 \qquad \text{and} \qquad \mathbb{E}\left[\mathrm{x}_i\epsilon_i\right]=\mathbb{E}\left[\mathrm{x}_i\mathbb{E}[\epsilon_i|\mathrm{x}_i]\right]=0$$

It follows that

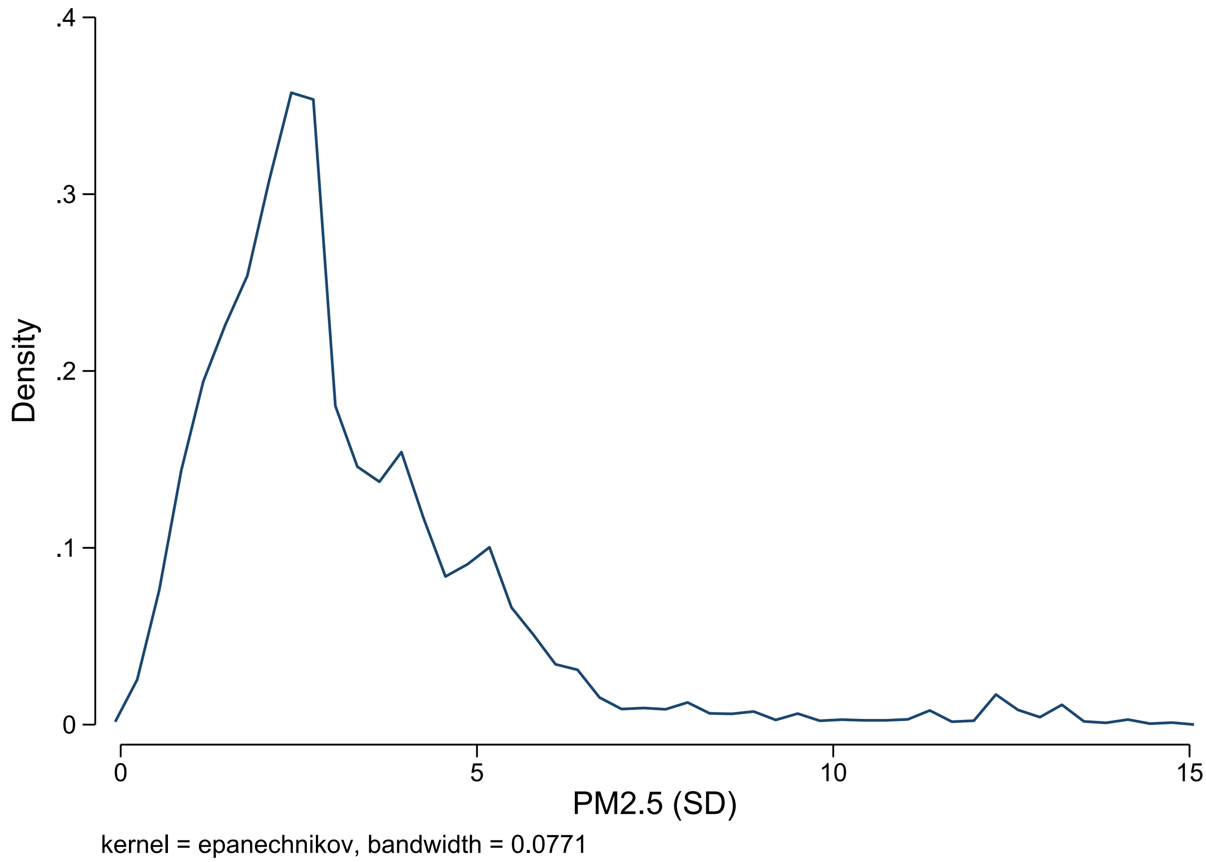
$$\text{plim}\frac{\hat{T}'M\boldsymbol{\epsilon}}{n}\xrightarrow{p}0$$

The second term in the RHS of equation (D8) converges in probability to zero. Hence,

$$\beta_{3s}\xrightarrow{p}\beta$$

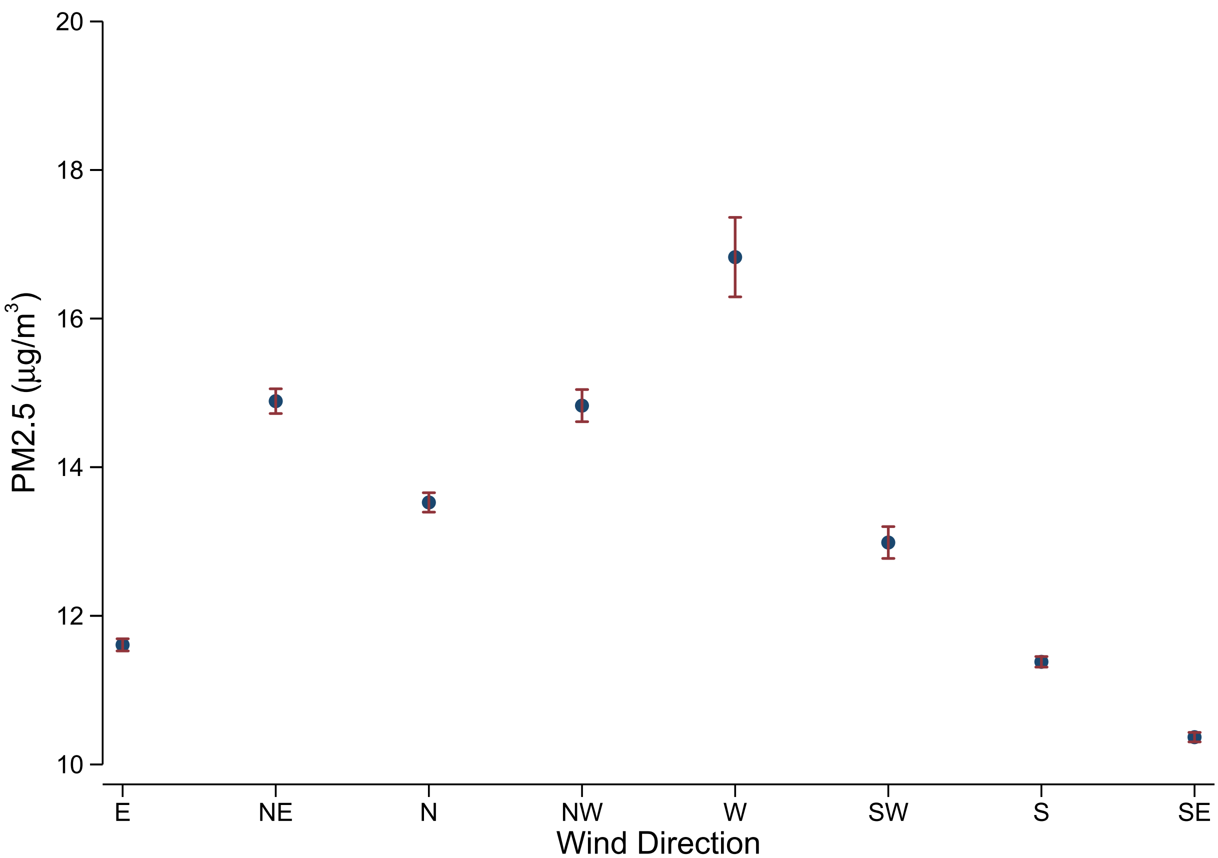
Appendix E. Additional Tables and Figures for Chapter III

Figure E1. $PM2.5$ SD Distribution



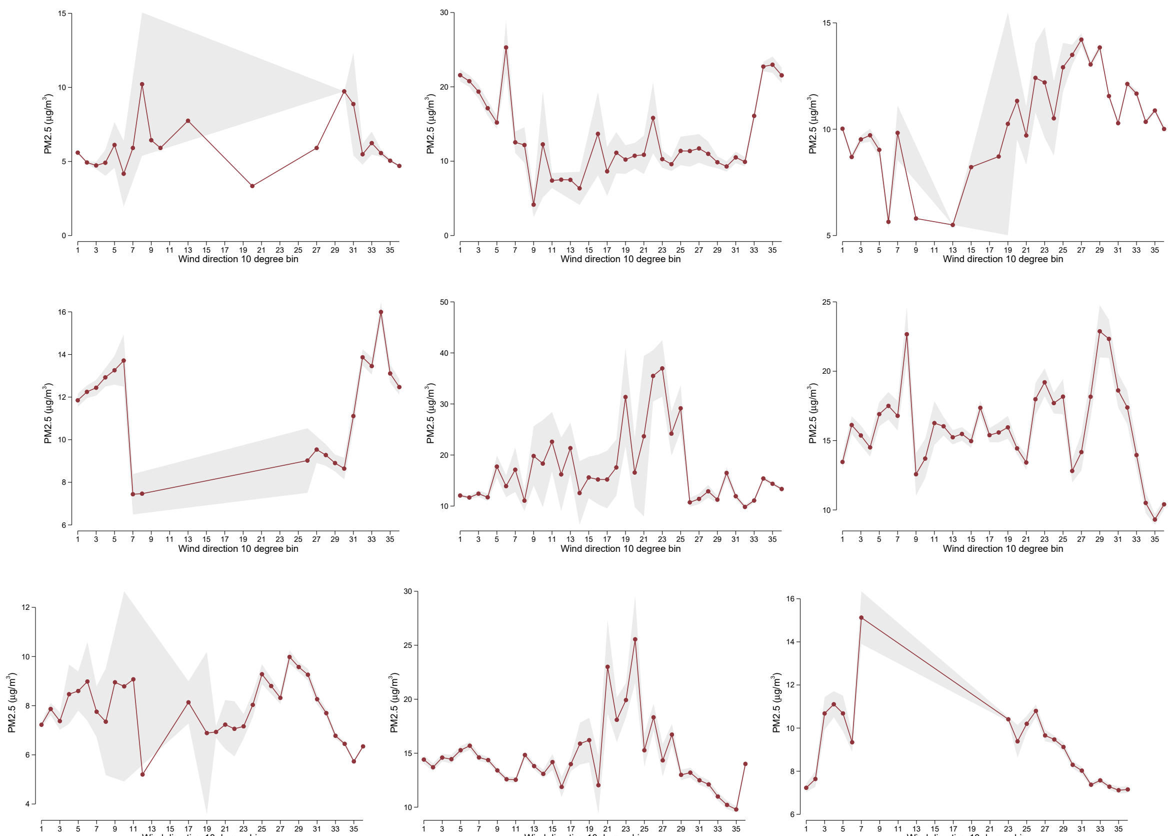
Note: This figure plots the distribution of within student standard deviation of $PM2.5$ concentration in Tanzania during our sample period.

Figure E2. $PM2.5$ Variation by Wind Direction



Note: Mean daily $PM2.5$ concentration by the octant the wind is blowing from along with the 95% confidence interval.

Figure E3. Instrument Variation



Note: This figure plots the estimates and the associated 95 percent confidence interval based on robust standard errors from equation (23) for each of the nine examination center clusters.

References

Abrevaya, Jason and Stephen G Donald, "A GMM approach for dealing with missing data on regressors," *Review of Economics and Statistics*, 2017, *99* (4), 657–662.

Afifi, AA and RM Elashoff, "Missing observations in multivariate statistics II. Point estimation in simple linear regression," *Journal of the American Statistical Association*, 1967, *62* (317), 10–29.

Aguilar-Gomez, Sandra, Holt Dwyer, Joshua Graff Zivin, and Matthew Neidell, "This Is Air: The "Nonhealth" Effects of Air Pollution," *Annual Review of Resource Economics*, 2022, *14* (1), null.

Aigner, Dennis J, "Regression with a binary independent variable subject to errors of observation," *Journal of Econometrics*, 1973, *1* (1), 49–59.

Ailshire, Jennifer A and Philippa Clarke, "Fine particulate matter air pollution and cognitive function among US older adults," *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 2015, *70* (2), 322–328.

Akmal, Maryam and Lant Pritchett, "Learning equity requires more than equality: Learning goals and achievement gaps between the rich and the poor in five developing countries," *International Journal of Educational Development*, 2021, *82*, 102350.

Al-Samarrai, Samer and Tessa Peasgood, "Educational attainments and household characteristics in Tanzania," *Economics of Education Review*, 1998, *17* (4), 395–417.

Almond, Douglas, Kenneth Y Chay, and David S Lee, "The costs of low birth weight," *The Quarterly Journal of Economics*, 2005, *120* (3), 1031–1083.

Amegah, A Kofi and Samuel Agyei-Mensah, "Urban air pollution in Sub-Saharan Africa: Time for action," *Environmental Pollution*, 2017, *220*, 738–743.

Anderson, Michael L, "As the Wind Blows: The Effects of Long-Term Exposure to Air Pollution on Mortality," *Journal of the European Economic Association*, 10 2019, *18* (4), 1886–1927.

Arceo, Eva, Rema Hanna, and Paulina Oliva, "Does the effect of pollution on infant mortality differ between developing and developed countries? Evidence from Mexico City," *The Economic Journal*, 2016, *126* (591), 257–280.

Austin, Wes, Garth Heutel, and Daniel Kreisman, "School bus emissions, student health and academic performance," *Economics of Education Review*, 2019, *70*, 109–126.

Balakrishnan, Uttara and Magda Tsaneva, "Air pollution and academic performance: Evidence from India," *World Development*, 2021, *146*, 105553.

Balietti, Anca, Souvik Datta, and Stefanija Veljanoska, "Air pollution and child development in India," *Journal of Environmental Economics and Management*, 2022, *113*, 102624.

Bau, Natalie and Jishnu Das, "Teacher Value Added in a Low-Income Country," *American Economic Journal: Economic Policy*, February 2020, *12* (1), 62–96.

Bedi, Arjun S., Marcos Y. Nakaguma, Brandon J. Restrepo, and Matthias Rieger, "Particle Pollution and Cognition: Evidence from Sensitive Cognitive Tests in Brazil," *Journal of the Association of Environmental and Resource Economists*, 2021, *8* (3), 443–474.

Bernaards, Coen A, Thomas R Belin, and Joseph L Schafer, "Robustness of a multivariate normal approximation for imputation of incomplete binary data," *Statistics in medicine*, 2007, *26* (6), 1368–1382.

Bharadwaj, Prashant, Matthew Gibson, Joshua Graff Zivin, and Christopher Neilson, "Gray matters: Fetal pollution exposure and human capital formation," *Journal of the Association of Environmental and Resource Economists*, 2017, *4* (2), 505–542.

Bishop, Kelly C, Jonathan D Ketcham, and Nicolai V Kuminoff, "Hazed and Confused: The Effect of Air Pollution on Dementia," Working Paper 24970, National Bureau of Economic Research August 2018.

Bitler, Marianne P, Janet Currie, and John Karl Scholz, "WIC eligibility and participation," *Journal of Human resources*, 2003, pp. 1139–1179.

Black, Dan A, Mark C Berger, and Frank A Scott, "Bounding parameter estimates with nonclassical measurement error," *Journal of the American Statistical Association*, 2000, *95* (451), 739–748.

Bollinger, Christopher R, "Bounding mean regressions when a binary regressor is mismeasured," *Journal of Econometrics*, 1996, *73* (2), 387–399.

_ and Martijn van Hasselt, "Bayesian moment-based inference in a regression model with misclassification error," *Journal of Econometrics*, 2017, *200* (2), 282–294.

Brachet, Tanguy, "Maternal smoking, misclassification, and infant health," *Misclassification, and Infant Health (September 16, 2008)*, 2008.

Brownstone, David and Camilla Kazimi, "Applying the bootstrap," *Research Paper. August*, 1998.

Burton, A and DG Altman, "Missing covariate data within cancer prognostic studies: a review of current reporting and proposed guidelines," *British journal of cancer*, 2004, *91* (1), 4–8.

Calderón-Garcidueñas, Lilian, Antonieta Mora-Tiscareño, Esperanza Ontiveros, Gilberto Gómez-Garza, Gerardo Barragán-Mejía, James Broadway, Susan Chapman, Gildardo Valencia-Salazar, Valerie Jewells, Robert R Maronpot et al., "Air pollution, cognitive deficits and brain abnormalities: a pilot study with children and dogs," *Brain and cognition*, 2008, *68* (2), 117–127.

Carneiro, Juliana, Matthew A. Cole, and Eric Strobl, "The Effects of Air Pollution on Students' Cognitive Performance: Evidence from Brazilian University Entrance Tests," *Journal of the Association of Environmental and Resource Economists*, 2021, *8* (6), 1051–1077.

Celhay, Pablo A, Bruce D Meyer, and Nikolas Mittag, "What Leads to Measurement Errors? Evidence from Reports of Program Participation in Three Surveys," Technical Report, National Bureau of Economic Research 2022.

Chen, Siyu, Chongshan Guo, and Xinfei Huang, "Air Pollution, Student Health, and School Absences: Evidence from China," *Journal of Environmental Economics and Management*, 2018, *92*, 465–497.

Chen, Xiaohong, Yingyao Hu, and Arthur Lewbel, "Nonparametric identification of regression models containing a misclassified dichotomous regressor without instruments," *Economics Letters*, 2008, *100* (3), 381–384.

_ , _ , and _ , "A note on the closed-form identification of regression models with a mismeasured binary regressor," *Statistics & probability letters*, 2008, *78* (12), 1473–1479.

Chmielewski, Anna K., "The Global Increase in the Socioeconomic Achievement Gap, 1964 to 2015," *American Sociological Review*, 2019, *84* (3), 517–544.

Cho, Hyunkuk, "Does Particulate Matter Affect Cognitive Performance?," *American Journal of Health Economics*, Forthcoming.

Coker, Eric and Samuel Kizito, "A narrative review on the human health effects of ambient air pollution in Sub-Saharan Africa: an urgent need for health effects studies," *International journal of environmental research and public health*, 2018, *15* (3), 427.

Coleman-Jensen, Alisha, Matthew P Rabbitt, Christian A Gregory, and Anita Singh, "Household Food Security in the United States in 2021," Technical Report 309 2022.

_ , _ , _ , and _ , "Statistical Supplement to Household Food Security in the United States in 2021," Technical Report 2022.

Courtemanche, Charles, Augustine Denteh, and Rusty Tchernis, "Estimating the associations between SNAP and food insecurity, obesity, and food purchases with imperfect administrative measures of participation," *Southern Economic Journal*, 2019, *86* (1), 202–228.

Currie, Janet, Eric A Hanushek, E. Megan Kahn, Matthew Neidell, and Steven G Rivkin, "Does Pollution Increase School Absences?," *The Review of Economics and Statistics*, 11 2009, *91* (4), 682–694.

Dagenais, Marcel G, "The use of incomplete observations in multiple regression analysis: A generalized least squares approach," *Journal of Econometrics*, 1973, *1* (4), 317–328.

Dardanoni, Valentino, Salvatore Modica, and Franco Peracchi, "Regression with imputed covariates: A generalized missing-indicator approach," *Journal of Econometrics*, 2011, *162* (2), 362–368.

Dempsey, Delia A and Neal L Benowitz, "Risks and benefits of nicotine to aid smoking cessation in pregnancy," *Drug safety*, 2001, *24*, 277–322.

Dercon, Stefan and Pramila Krishnan, "In Sickness and in Health: Risk Sharing within Households in Rural Ethiopia," *Journal of Political Economy*, 2000, *108* (4), 688–727.

Deryugina, Tatyana, Garth Heutel, Nolan H. Miller, David Molitor, and Julian Reif, "The Mortality and Medical Costs of Air Pollution: Evidence from Changes in Wind Direction," *American Economic Review*, December 2019, *109* (12), 4178–4219.

DiTraglia, Francis J and Camilo García-Jimeno, "Mis-classified, binary, endogenous regressors: Identification and inference," Technical Report, National Bureau of Economic Research 2017.

Dockery, Douglas and Arden Pope, "Epidemiology of acute health effects: summary of time-series studies," 1996.

Ebenstein, Avraham, Victor Lavy, and Sefi Roth, "The Long-Run Economic Consequences of High-Stakes Examinations: Evidence from Transitory Variation in Pollution," *American Economic Journal: Applied Economics*, October 2016, *8* (4), 36–65.

Efron, B., "Bootstrap Methods: Another Look at the Jackknife," *The Annals of Statistics*, 1979, *7* (1), 1–26.

Efron, Bradley and Robert J Tibshirani, *An introduction to the bootstrap*, Chapman and Hall/CRC, 1994.

Einarson, Adrienne and Sara Riordan, "Smoking in pregnancy and lactation: a review of risks and cessation strategies," *European journal of clinical pharmacology*, 2009, *65*, 325–330.

Evans, William N and Jeanne S Ringel, "Can higher cigarette taxes improve birth outcomes?," *Journal of public Economics*, 1999, *72* (1), 135–154.

Fertig, Angela R, "Selection and the effect of prenatal smoking," *Health Economics*, 2010, *19* (2), 209–226.

Fowlie, Meredith, Edward Rubin, and Reed Walker, "Bringing Satellite-Based Air Quality Estimates Down to Earth," *AEA Papers and Proceedings*, May 2019, *109*, 283–88.

Fox, Liana E and Charles Hokayem, "Using WIC Administrative Data to Evaluate the Supplemental Poverty Measure," 2022.

Frazis, Harley and Mark A Loewenstein, "Estimating linear regressions with mismeasured, possibly endogenous, binary explanatory variables," *Journal of Econometrics*, 2003, *117* (1), 151–178.

Galimard, Jacques-Emmanuel, Sylvie Chevret, Emmanuel Curis, and Matthieu Resche-Rigon, "Heckman imputation models for binary or continuous MNAR outcomes and MAR predictors," *BMC medical research methodology*, 2018, *18* (1), 1–13.

Garber, Steven and Steven Klepper, "Extending the classical normal errors-in-variables model," *Econometrica: Journal of the Econometric Society*, 1980, pp. 1541–1546.

Gilraine, Michael and Angela Zheng, "Air Pollution and Student Performance in the U.S.," Working Paper 30061, National Bureau of Economic Research May 2022.

Gleixner, Stephanie, Teferi Demissie, and Gulilat Tefera Diro, "Did ERA5 Improve Temperature and Precipitation Reanalysis over East Africa?," *Atmosphere*, 2020, *11* (9).

Graff Zivin, Joshua, Tong Liu, Yingquan Song, Qu Tang, and Peng Zhang, "The unintended impacts of agricultural fires: Human capital in China," *Journal of Development Economics*, 2020, *147*, 102560.

Greene, William H, "Econometric analysis, 71e," *Stern School of Business, New York University*, 2012.

Gruber, Jonathan and Botond Köszegi, "Is addiction "rational"? Theory and evidence," *The Quarterly Journal of Economics*, 2001, *116* (4), 1261–1303.

Gundersen, Craig and James P Ziliak, "Food insecurity research in the United States: Where we have been and where we need to go," *Applied Economic Perspectives and Policy*, 2018, *40* (1), 119–135.

_ , Brent Kreider, and John Pepper, "The economics of food insecurity in the United States," *Applied Economic Perspectives and Policy*, 2011, *33* (3), 281–303.

Hadley, Craig, David Lindstrom, Fasil Tessema, and Tefara Belachew, "Gender bias in the food insecurity experience of Ethiopian adolescents," *Social Science & Medicine*, 2008, *66* (2), 427–438.

Hausman, Jerry A, Jason Abrevaya, and Fiona M Scott-Morton, "Misclassification of the dependent variable in a discrete-response setting," *Journal of econometrics*, 1998, *87* (2), 239–269.

Heissel, Jennifer A., Claudia Persico, and David Simon, "Does Pollution Drive Achievement? The Effect of Traffic Pollution on Academic Performance," *Journal of Human Resources*, 2022, *57* (3), 747–776.

Heitjan, Daniel F and Srabashi Basu, "Distinguishing "missing at random" and "missing completely at random"," *The American Statistician*, 1996, *50* (3), 207–213.

Horowitz, Joel L, "The bootstrap," in "Handbook of econometrics," Vol. 5, Elsevier, 2001, pp. 3159–3228.

Horton, Nicholas J and Ken P Kleinman, "Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models," *The American Statistician*, 2007, *61* (1), 79–90.

Hu, Yingyao, Ji-Liang Shiu, and Tiemen Woutersen, "Identification and estimation of single-index models with measurement error and endogeneity," *The Econometrics Journal*, 2015, *18* (3), 347–362.

—, —, and —, "Identification in nonseparable models with measurement errors and endogeneity," *Economics Letters*, 2016, *144*, 33–36.

III, C Arden Pope and Douglas W Dockery, "Health effects of fine particulate air pollution: lines that connect," *Journal of the air & waste management association*, 2006, *56* (6), 709–742.

Inness, A., M. Ades, A. Agustí-Panareda, J. Barré, A. Benedictow, A.-M. Blechschmidt, J. J. Dominguez, R. Engelen, H. Eskes, J. Flemming, V. Huijnen, L. Jones, Z. Kipling, S. Massart, M. Parrington, V.-H. Peuch, M. Razinger, S. Remy, M. Schulz, and M. Suttie, "The CAMS reanalysis of atmospheric composition," *Atmospheric Chemistry and Physics*, 2019, *19* (6), 3515–3556.

Jackson, Dan, Ian R White, Dan Mason, and Stephen Sutton, "A general method for handling missing binary outcome data in randomized controlled trials," *Addiction*, 2014, *109* (12), 1986–1993.

Jensen, Helen H, Brent Kreider, and Oleksandr Zhylyevskyy, "Investigating treatment effects of participating jointly in SNAP and WIC when the treatment is validated only for SNAP," *Southern Economic Journal*, 2019, *86* (1), 124–155.

Jones, Michael P, "Indicator and stratification methods for missing explanatory variables in multiple linear regression," *Journal of the American statistical association*, 1996, *91* (433), 222–230.

Kampa, Marilena and Elias Castanas, "Human health effects of air pollution," *Environmental pollution*, 2008, *151* (2), 362–367.

Kane, Thomas J, Cecilia E Rouse, and Douglas O Staiger, "Estimating returns to schooling when schooling is misreported," 1999.

Klasen, Stephan, "Nutrition, health and mortality in sub□Saharan Africa: Is there a gender bias?," *The Journal of Development Studies*, 1996, *32* (6), 913–932.

Komisarow, Sarah and Emily L Pakhtigian, "Are power plant closures a breath of fresh air? Local air quality and school absences," *Journal of Environmental Economics and Management*, 2022, *112*, 102569.

Kreider, Brent, "Regression coefficient identification decay in the presence of infrequent classification errors," *The Review of Economics and Statistics*, 2010, *92* (4), 1017–1023.

—, John V Pepper, Craig Gundersen, and Dean Jolliffe, "Identifying the effects of SNAP (food stamps) on child health outcomes when participation is endogenous and misreported," *Journal of the American Statistical Association*, 2012, *107* (499), 958–975.

Lee, David S, Justin McCrary, Marcelo J Moreira, and Jack R Porter, "Valid t-ratio Inference for IV," Working Paper 29124, National Bureau of Economic Research August 2021.

Lewbel, Arthur, "Estimation of average treatment effects with misclassification," *Econometrica*, 2007, *75* (2), 537–551.

Lien, Diana S and William N Evans, "Estimating the impact of large cigarette tax hikes: the case of maternal smoking and infant birth weight," *Journal of Human resources*, 2005, *40* (2), 373–392.

Little, Roderick JA, "Regression with missing X's: a review," *Journal of the American statistical association*, 1992, *87* (420), 1227–1237.

_ and Donald B Rubin, "Statistical analysis with missing data," *Statistical analysis with missing data*, 2002.

Loftus, Christine T., Marnie F. Hazlehurst, Adam A. Szpiro, Yu Ni, Frances A. Tylavsky, Nicole R. Bush, Sheela Sathyanarayana, Kecia N. Carroll, Catherine J. Karr, and Kaja Z. LeWinn, "Prenatal air pollution and childhood IQ: Preliminary evidence of effect modification by folate," *Environmental Research*, 2019, *176*, 108505.

Lushniak, Boris D, Jonathan M Samet, Terry F Pechacek, Leslie A Norman, and Peter A Taylor, "The health consequences of smoking—50 years of progress: A report of the surgeon general," 2014.

Mahajan, Aprajit, "Identification and estimation of regression models with misclassification," *Econometrica*, 2006, *74* (3), 631–665.

Maity, Arnab Kumar, Vivek Pradhan, and Ujjwal Das, "Bias reduction in logistic regression with missing responses when the missing data mechanism is nonignorable," *The American Statistician*, 2019.

Mancino, Lisa, Jessica E Todd, and Benjamin Scharadin, "Usda's national household food acquisition and purchase survey: Methodology for imputing missing quantities to calculate healthy eating index-2010 scores and sort foods into ers food groups," Technical Report 2018.

Marcotte, Dave E, "Something in the air? Air quality and children's educational outcomes," *Economics of Education Review*, 2017, *56*, 141–151.

McDonough, Ian K and Daniel L Millimet, "Missing data, imputation, and endogeneity," *Journal of Econometrics*, 2017, *199* (2), 141–155.

Meyer, Bruce D, Nikolas Mittag, and Robert M Goerge, "Errors in survey reporting and imputation and their effects on estimates of food stamp program participation," *Journal of Human Resources*, 2022, *57* (5), 1605–1644.

_ , Wallace KC Mok, and James X Sullivan, "Household surveys in crisis," *Journal of Economic Perspectives*, 2015, *29* (4), 199–226.

Mills, Nicholas L, Ken Donaldson, Paddy W Hadoke, Nicholas A Boon, William MacNee, Flemming R Cassee, Thomas Sandström, Anders Blomberg, and David E Newby, "Adverse cardiovascular effects of air pollution," *Nature clinical practice Cardiovascular medicine*, 2009, *6* (1), 36–44.

MoEST, “EDUCATION SECTOR PERFORMANCE REPORT (2018/2019),” Technical Report, Minister for Education, Science and Technology 2019.

Nauze, Andrea La and Edson R Severnini, “Air Pollution and Adult Cognition: Evidence from Brain Training,” Working Paper 28785, National Bureau of Economic Research May 2021.

Nguimkeu, Pierre, Augustine Denteh, and Rusty Tchernis, “On the estimation of treatment effects with endogenous misreporting,” *Journal of econometrics*, 2019, *208* (2), 487–506.

_ , Robert Rosenman, and Vidhura Tennekoon, “Regression with a misclassified binary regressor: Correcting for the hidden bias,” 2021.

noz Sabater, J. Mu E. Dutra, A. Agustí-Panareda, C. Albergel, G. Arduini, G. Balsamo, S. Boussetta, M. Choulga, S. Harrigan, H. Hersbach, B. Martens, D. G. Miralles, M. Piles, N. J. Rodríguez-Fernández, E. Zsoter, C. Buontempo, and J.-N. Thépaut, “ERA5-Land: a state-of-the-art global reanalysis dataset for land applications,” *Earth System Science Data*, 2021, *13* (9), 4349–4383.

Oberdörster, G, Z Sharp, V Atudorei, A Elder, R Gelein, W Kreyling, and C Cox, “Translocation of inhaled ultrafine particles to the brain,” *Inhalation toxicology*, 2004, *16* (6-7), 437–445.

Oliveira, Victor, “The Food Assistance Landscape: FY 2017 Annual Report,” Technical Report EIB-190, U.S. Department of Agriculture, Economic Research Service March 2018.

Persico, Claudia L and Joanna Venator, “The effects of local industrial pollution on students and schools,” *Journal of Human Resources*, 2021, *56* (2), 406–445.

Peters, Annette, Bellina Veronesi, Lilian Calderón-Garcidueñas, Peter Gehr, Lung Chi Chen, Marianne Geiser, William Reed, Barbara Rothen-Rutishauser, Samuel Schürch, and Holger Schulz, “Translocation and potential neurological effects of fine and ultrafine particles a critical update,” *Particle and fibre toxicology*, 2006, *3* (1), 1–13.

Poirier, Dale J, “Partial observability in bivariate probit models,” *Journal of econometrics*, 1980, *12* (2), 209–217.

Ridolo, Erminia, Cristoforo Incorvaia, Irene Martignago, Marco Caminati, Giorgio Walter Canonica, and Gianenrico Senna, “Sex in Respiratory and Skin Allergies,” *Clinical Reviews in Allergy & Immunology*, Jun 2019, *56* (3), 322–332.

Rubin, Donald B, “Inference and missing data,” *Biometrika*, 1976, (63), 581–592.

_ , “An overview of multiple imputation,” in “Proceedings of the survey research methods section of the American statistical association,” Vol. 79 Citeseer 1988, p. 84.

_ , *Multiple imputation for nonresponse in surveys*, Vol. 81, John Wiley & Sons, 2004.

Savoca, Elizabeth, “Measurement errors in binary regressors: an application to measuring the effects of specific psychiatric diseases on earnings,” *Health Services and Outcomes Research Methodology*, 2000, *1*, 149–164.

Scates, Sara S, Timothy P Finn, Janna Wisniewski, David Dadi, Renata Mandike, Mwinyi Khamis, George Greer, Naomi Serbantez, Sylvester Segbaya, Prince Owusu et al., "Costs of insecticide-treated bed net distribution systems in sub-Saharan Africa," *Malaria Journal*, 2020, *19*, 1–18.

Schady, Norbert R. Fiszbein Ariel, "Conditional Cash Transfers Reducing Present and Future Poverty," Technical Report 978-0-8213-7352-1, World Bank February 2009.

Schafer, Joseph L, "Analysis of incomplete multivariate data," *Chapman &Hall/CRC*, 1997.

Schikowski, Tamara and Hicran Altuğ, "The role of air pollution in cognitive impairment and decline," *Neurochemistry International*, 2020, *136*, 104708.

Settele, Sonja and Reyn Van Ewijk, "Can cigarette taxes during pregnancy mitigate the intergenerational transmission of socioeconomic status?," *Labour Economics*, 2018, *55*, 130–148.

Simon, David, "Does early life exposure to cigarette smoke permanently harm childhood welfare? Evidence from cigarette tax hikes," *American Economic Journal: Applied Economics*, 2016, *8* (4), 128–159.

Singh, Tejendra Pratap, "Beyond The Haze: Air Pollution and Student Absenteeism - Evidence from India," OSF Preprints pcva2, Center for Open Science May 2022.

Ssenyunzi, Richard Cliffe, Bosco Oruru, Florence Mutonyi D'ujanga, Eugenio Realini, Stefano Barindelli, Giulio Tagliaferro, Axel von Engeln, and Nick van de Giesen, "Performance of ERA5 data in retrieving Precipitable Water Vapour over East African tropical region," *Advances in Space Research*, 2020, *65* (8), 1877–1893.

Subasi, Munevver Mine, Ersoy Subasi, Martin Anthony, and Peter L Hammer, "A new imputation method for incomplete binary data," *Discrete applied mathematics*, 2011, *159* (10), 1040–1047.

The Examinations Regulations, "The National Examinations Council Act Cap. 107 , The Examinations Regulations, 2015.," 2015.

Ura, Takuya, "Heterogeneous treatment effects with mismeasured endogenous treatment," *Quantitative Economics*, 2018, *9* (3), 1335–1370.

van Donkelaar, Aaron, Melanie S. Hammer, Liam Bindle, Michael Brauer, Jeffery R. Brook, Michael J. Garay, N. Christina Hsu, Olga V. Kalashnikova, Ralph A. Kahn, Colin Lee, Robert C. Levy, Alexei Lyapustin, Andrew M. Sayer, and Randall V. Martin, "Monthly Global Estimates of Fine Particulate Matter and Their Uncertainty," *Environmental Science & Technology*, 2021, *55* (22), 15287–15300.

van Hasselt, Martijn and Christopher R Bollinger, "Binary misclassification and identification in regression models," *Economics Letters*, 2012, *115* (1), 81–84.

White, Ian R, Rhian Daniel, and Patrick Royston, "Avoiding bias due to perfect prediction in multiple imputation of incomplete categorical variables," *Computational statistics & data analysis*, 2010, *54* (10), 2267–2275.

Wilker, Elissa H, Sarah R Preis, Alexa S Beiser, Philip A Wolf, Rhoda Au, Itai Kloog, Wenyuan Li, Joel Schwartz, Petros Koutrakis, Charles DeCarli et al., "Long-term exposure to fine particulate matter, residential proximity to major roads and measures of brain structure," *Stroke*, 2015, *46* (5), 1161–1166.

Yusuph, Mashala, "The Impact of the Implementation of Free Education Policy on Secondary Education in Tanzania," *International Journal of Academic Multidisciplinary Research*, January 2019, *3* (1), 6–14.

Vita

Erica Mtenga received her Ph.D. in economics from Georgia State University in August 2025. Before then, Erica graduated in November 2015 from the University of Dar es Salaam with a B.A. in economics and statistics. Erica also received an M.A. in economics from the University of Dar es Salaam in November 2017 and Georgia State University in August 2023.