Finite-population inference via GMM estimator

Haruo Kakehi and Taisuke Otsu

March 5, 2024

KEO Discussion Paper No. 178

KEO Discussion Paper No. 178

FINITE-POPULATION INFERENCE VIA GMM ESTIMATOR

HARUO KAKEHI AND TAISUKE OTSU

ABSTRACT. Since the seminal works by Abadie *et al.* (2014, 2020), there has been considerable attention on finite-population inference for various econometric problems. This paper extends the finite-population asymptotic approach to the generalized method of moments (GMM) estimator for overidentified moment condition models. A motivating setup is the situation where researchers have auxiliary information on some population moments. Under the finite-population asymptotics, we study asymptotic properties of the GMM estimator, propose asymptotically conservative variance estimators, and discuss how to select the GMM weight matrix. A simulation study based on real data for entrepreneurship training and incubation programs illustrate usefulness of the proposed method.

1. INTRODUCTION

Since the seminal works by Abadie *et al.* (2014, 2020), there has been growing interest in inference methods under finite-population setups accounting for design-based uncertainty. The design-based perspective for investigating econometric or statistical methods has been prevalent in randomized experiments (e.g., Neyman, 1923; Rosenbaum, 2002; Freedman, 2008a, 2008b). However, the literature does not consider the sampling-based uncertainty deriving from not observing the entire population since it is common to assume that random assignment is the only source of uncertainty in an experimental setting. On the other hand, extensive statistical literature exists on finite-population asymptotics, taking into account of the sampling variation (see, Práŝková and Sen, 2009 for an overview) although this body of work omits consideration on the design-based uncertainty. Abadie *et al.* (2020) developed an alternative inference framework in observational study settings by incorporating both design and sampling-based uncertainty. While their framework is restricted to the case of linear regression, the recent literature considers design-based uncertainty in different settings; M-estimators (Xu, 2021), spatial correlation (Xu and Wooldridge, 2022), difference-in-differences (Rambachan and Roth, 2022), and staggered difference-in-differences (Athey and Imbens, 2022; Roth and Sant'Anna, 2023). Our paper closely relates to the work by Xu (2021), which extends Abadie *et al.* (2020) to the M-estimator setting.

In this paper, we extend the above existing results on finite-population inference to the situations where estimands of interest are defined by overidentified moment conditions. In conventional empirical economic analyses using the infinite-population asymptotics, overidentified moment condition models are ubiquitous, and there is rich literature on applied and theoretical econometric analyses for these models typically using the generalized method of moments

(GMM) (see, e.g., Hall, 2005, for an overview). Therefore, it is of substantial interest to develop the GMM theory for finite-population inference problems. One motivating example is the situation where the researcher has auxiliary information on the population moments as investigated in Imbens and Lancaster (1994). Under the finite-population asymptotic framework, the proportion of the sample size to the population size is non-negligible and it may be plausible that researchers have access to some population moments. Examples include large-scale experiments where sample representativeness is important (Muralidharan and Niehaus, 2017; Duflo and Banerjee, 2017) and the Integrated Public Use Microdata Series (IPUMS) data, which is the 10% sample of the U.S. Census. In these cases, researchers may incorporate moments based on the entire states or counties to improve point estimators and associated inference. Furthermore, the finite-population asymptotic analysis of the GMM estimator and development of feasible inference methods are open issues in the literature. In contrast to M-estimation problems as studied in Xu (2021), the GMM requires a weight matrix, and currently there is no guidance on its choice under the finite-population asymptotics.

This paper studies asymptotic properties of the GMM estimator for overidentified moment condition models under the finite-population asymptotics. We derive the consistency and asymptotic normality of the GMM estimator. In particular, we find that its finite-population asymptotic variance takes a different form from the conventional infinite-population asymptotic variance, which is an analogous finding in just-identified modes (e.g., Abadie *et al.,* 2020; Xu, 2021). Since our asymptotic variance for the GMM estimator involves a component which is not consistently estimable, we propose two asymptotically conservative variance estimators: one is the conventional variance estimator and the other is an adapted version of Abadie *et al.* (2020)'s variance estimator to the GMM context. Furthermore, we discuss the choice of the GMM weight matrix under the finite-population asymptotic framework and suggest a feasible data-dependent weight, where the associated conservative variance estimator shows a desirable property. These theoretical findings are illustrated by a empirically motivated simulation study by using real data on entrepreneurship training and incubation programs in North American undergraduates.

The rest of the paper is organized as follows. Section 2 presents our main results. After introducing our basic setup and the GMM estimator in Section 2.1, Section 2.2 presents asymptotic properties of the GMM estimator under the finite-population asymptotics and studies estimation of the asymptotic variance, and we discuss the choice of the GMM weight matrix in Section 2.3. Section 3 illustrates the proposed inference method by a simulation study based on the real data for entrepreneurial activities of minority groups, where auxiliary information on some population moments are available as in Imbens and Lancaster (1994).

2. Main results

2.1. **Setup and estimator.** We first introduce our basic setup based on Abadie *et al.* (2020) and Xu (2021). For each unit $i = 1, \dots, M$ with population size M, consider the population $\{X_i, z_i, Y_i\}_{i=1}^{M}$, where X_i is a vector of assignment variables, z_i is a vector of i's attributions, and Y_i is a vector of outcome variables. Our interest is in the effect of X_i rather than z_i. For example, in empirical analysis, X_i refers to a treatment status of an experiment or indicator for a policy intervention, while z_i includes e.g. age, gender, and socioeconomic status of individuals. Throughout the paper, we assume that z_i is non-random, and the outcomes are written as $Y_i = y_i(X_i)$ for a potential outcome function $y_i(\cdot)$. Although these variables depend on the population size M to conduct asymptotic analysis for $M \to \infty$, we suppress such dependence to simplify the presentation. From the finite population $\{X_i, z_i, Y_i\}_{i=1}^M$, we observe a random sample by Bernoulli sampling. Let R_i be a Bernoulli random variable, which equals to one if i is sampled, and zero otherwise. Let $N = \sum_{i=1}^M R_i$ be the sample size, which is random.

This paper is concerned with the situation where the estimand of interest is defined as a solution of overidentified moment restrictions. Let $g_i(X_i, \theta) := g(X_i, z_i, y_i(X_i), \theta)$ with a k-dimensional vector of moment functions g and a p-dimensional vector of parameters. The estimand of interest θ_M^* is defined as a unique solution of

$$\frac{1}{M}\sum_{i=1}^{M}\mathbb{E}[g_i(X_i, \theta_M^*)] = 0, \tag{2.1}$$

for $k > p$, i.e., the moment condition (2.1) is overidentified. Note that the expectation $\mathbb{E}[\cdot]$ is taken with respect to the assignment variables X_i. It should be noted that the existing papers such as Abadie *et al.* (2020) and Xu (2021) do not cover such a setup.

As an estimator of θ_M^*, this paper focuses on the GMM estimator

$$\hat{\theta}_N(W_N) = \arg\min_{\theta\in\Theta}\left\{\frac{1}{N}\sum_{i=1}^{M}R_i g_i(X_i,\theta)\right\}' W_N \left\{\frac{1}{N}\sum_{i=1}^{M}R_i g_i(X_i,\theta)\right\},$$

where Θ is parameter space of θ_M^* and W_N is a $k \times k$ weight matrix. The choice of W_N is discussed in Section 2.3 below.

2.2. **Finite-population asymptotic theory.** We now study large sample properties of the GMM estimator $\hat{\theta}_N(W_N)$ in our asymptotic framework. We impose the following assumptions.

Assumption.

1. $\{X_i\}_{i=1}^M$ *is independent but not necessarily identically distributed.* $\{R_i\}_{i=1}^M$ *is independent and identically distributed sample of Bernoulli random variables with* $\rho_M = \mathbb{P}(R_i = 1)$ *satisfying* $\rho_M \to \rho \in [0,1]$ *as* $M \to \infty$. *Furthermore,* $\{X_i\}_{i=1}^M$ *and* $\{R_i\}_{i=1}^M$ *are independent.*
2. Θ *is compact.* W_N *converges in probability to a positive definite matrix* W. $\lim_{M\to\infty}\frac{1}{M}\sum_{i=1}^M \mathbb{E}[g_i(X_i,\theta)] = 0$ *is uniquely satisfied at* $\theta = \theta^* := \lim_{M\to\infty}\theta_M^*$. $g_i(x,\theta)$ *is continuous at each* $\theta \in \Theta$ *for almost every* x, *and* $\sup_{i,M}\mathbb{E}[\sup_{\theta\in\Theta}||g_i(X_i,\theta)||^4] < \infty$. *There exists functions* $h_1(\cdot)$ *and* $b_{1i}(\cdot)$ *such that* $\lim_{u\to 0}h_1(u) = 0$, $\sup_{i,M}\mathbb{E}[b_{1i}(X_i)] < \infty$, *and* $||g_i(X_i,\theta) - g_i(X_i,\theta_1)|| \le b_{1i}(X_i)h_1(||\theta-\theta_1||)$ *for each* $\theta, \theta_1 \in \Theta$.
3. $\theta^* \in \mathrm{int}(\Theta)$. $g_i(x,\theta)$ *is continuously differentiable on* $\mathrm{int}(\Theta)$ *for almost every* x, *and* $G := \lim_{M\to\infty}\frac{1}{M}\sum_{i=1}^M \mathbb{E}\left[\frac{\partial g_i(X_i,\theta_M^*)}{\partial\theta'}\right]$ *has full column rank. There exists functions* $h_2(\cdot)$ *and* $b_{2i}(\cdot)$ *and a neighborhood* $\mathcal{N}$ *around* θ^* *such that* $\sup_{i,M}\mathbb{E}[\sup_{\theta\in\mathcal{N}}||\partial g_i(X_i,\theta)/\partial\theta'||^2] < \infty$, $\lim_{u\to 0}h_2(u) = 0$, $\sup_{i,M}\mathbb{E}[b_{2i}(X_i)] < \infty$, *and* $||\partial g_i(X_i,\theta)/\partial\theta' - \partial g_i(X_i,\theta_1)/\partial\theta'|| \le b_{2i}(X_i)h_2(||\theta-\theta_1||)$ *for each* $\theta, \theta_1 \in \mathcal{N}$.

Assumption (1) is on the sampling framework, which is also employed by Abadie *et al.* (2020) and Xu (2021). This implies the sample size $N = \sum_{i=1}^{M} R_i$ is random and its expectation $\mathbb{E}[N] = M\rho_M$ diverges at the same rate as $M \to \infty$. Assumptions (2) and (3) collect regularity conditions on the weight matrix W_N and the moment function g_i. These are natural adaptations of the conventional GMM theory to our finite-population setup. Assumption (2) is used to derive the consistency of the GMM estimator, and Assumption (3) contains additional conditions to establish asymptotic normality.

Under the above assumptions, the asymptotic properties of the GMM estimator $\hat{\theta}_N(W_N)$ are obtained as follows.

Theorem 1.

(1) *Under Assumptions (1)-(2), it holds* $\hat{\theta}_N(W_N) - \theta_M^* \xrightarrow{p} 0$.

(2) *Under Assumptions (1)-(3), it holds*

$$\sqrt{N}(\hat{\theta}_N(W_N) - \theta_M^*) \xrightarrow{d} N(0, V_{\text{GMM}}(W)), \tag{2.2}$$

where

$$\begin{aligned} V_{\text{GMM}}(W) &= (G'WG)^{-1}G'W(\Omega - \rho\Delta)WG(G'WG)^{-1}, \\ \Omega &= \lim_{M\to\infty} \frac{1}{M}\sum_{i=1}^{M} \mathbb{E}[g_i(X_i, \theta_M^*)g_i(X_i, \theta_M^*)'], \\ \Delta &= \lim_{M\to\infty} \frac{1}{M}\sum_{i=1}^{M} \mathbb{E}[g_i(X_i, \theta_M^*)]\mathbb{E}[g_i(X_i, \theta_M^*)]'. \end{aligned}$$

Theorem 1 (1) says that the GMM estimator $\hat{\theta}_N(W_N)$ is consistent for the population parameter θ_M^*, and Theorem 1 (2) derives its asymptotic distribution. Compared to the conventional infinite-population asymptotics, the main difference is presence of the additional term "$\rho\Delta$" in the asymptotic variance. Letting $V_{\text{C}}(W) = (G'WG)^{-1}G'W\Omega WG(G'WG)^{-1}$ be the asymptotic variance of the GMM estimator under the conventional infinite-population asymptotics, and $V_{\text{A}}(W) = \rho(G'WG)^{-1}G'W\Delta WG(G'WG)^{-1}$ be an additional component, the finite-population asymptotic variance can be written as $V_{\text{GMM}}(W) = V_{\text{C}}(W) - V_{\text{A}}(W)$. Since Δ is positive semi-definite, $V_{\text{A}}(W)$ is also positive semi-definite and $V_{\text{GMM}}(W)$ is always smaller than the conventional variance $V_{\text{C}}(W)$ in the matrix sense (denoted by $V_{\text{GMM}}(W) \leq_{\text{pd}} V_{\text{C}}(W)$). Although $V_{\text{C}}(W)$ can be consistently estimated (as shown below), the component Δ and thus the variance $V_{\text{GMM}}(W)$ cannot be consistently estimable in general.

As in Abadie *et al.* (2020) and Xu (2021), we propose conservative estimators for $V_{\text{GMM}}(W)$. The first variance estimator is a consistent estimator of the conventional variance $V_{\text{C}}(W)$, that is

$$\hat{V}_{\text{C}}(W_N) = (\hat{G}'W_N\hat{G})^{-1}\hat{G}'W_N\hat{\Omega}W_N\hat{G}(\hat{G}'W_N\hat{G})^{-1},$$

where $\hat{G} = \frac{1}{N}\sum_{i=1}^{M} R_i \partial g_i(X_i, \hat{\theta}_N(W_N))/\partial\theta'$ and $\hat{\Omega} = \frac{1}{N}\sum_{i=1}^{M} R_i g_i(X_i, \hat{\theta}_N(W_N)) g_i(X_i, \hat{\theta}_N(W_N))'$. The second variance estimator is constructed by estimating a lower bound for Δ, that is

$$\hat{\Delta}_Z = \frac{1}{N}\sum_{i=1}^{M} R_i \hat{P}' z_i z_i' \hat{P},$$

where $\hat{P} = (\sum_{i=1}^{M} R_i z_i z_i')^{-1}(\sum_{i=1}^{M} R_i z_i g_i(X_i, \hat{\theta}_N(W_N)))$. Then the second variance estimator is

$$\hat{V}_Z(W_N) = (\hat{G}' W_N \hat{G})^{-1} \hat{G}' W_N (\hat{\Omega} - \rho\hat{\Delta}_Z) W_N \hat{G} (\hat{G}' W_N \hat{G})^{-1}.$$

The asymptotic properties of these variance estimators are obtained as follows.

Theorem 2.

(1) *Under Assumptions (1)-(3),* $\hat{V}_{\mathrm{C}}(W_N) \xrightarrow{p} V_{\mathrm{C}}(W)$.

(2) *In addition to Assumptions (1)-(3), assume that* $\lim_{M\to\infty} \frac{1}{M}\sum_{i=1}^{M} \mathbb{E}[g_i(X_i, \theta_M^*)] z_i'$ *exists,* $\lim_{M\to\infty} \frac{1}{M}\sum_{i=1}^{M} z_i z_i'$ *exists and non-singular, and* $\sup_{i,M} ||z_i|| < \infty$. *Then* $\hat{\Delta}_Z$ *converges in probability to a positive semi-definite matrix* Δ_Z *such that* $\Delta_Z \leq_{\mathrm{pd}} \Delta$.

(3) $V_{\mathrm{GMM}}(W) \leq_{\mathrm{pd}} V_Z(W) \leq_{\mathrm{pd}} V_{\mathrm{C}}(W)$, *where* $V_Z(W) = (G'WG)^{-1}G'W(\Omega - \rho\Delta_Z)WG(G'WG)^{-1}$.

Since the proof of this theorem is similar to the ones in Xu (2021, Theorems 2.2 and 3.1), it is omitted. Theorem 2 (1) says that the conventional variance estimator $\hat{V}_{\mathrm{C}}(W_N)$ is still consistent for the variance component $V_{\mathrm{C}}(W)$ under the finite-population asymptotics. Theorem 2 (2) guarantees conservativeness of the second variance estimator $\hat{V}_Z(W_N)$ for $V_{\mathrm{GMM}}(W)$. Theorem 2 (3) clarifies the relationships of the limits of the variance estimators. Although we cannot consistently estimate the asymptotic variance $V_{\mathrm{GMM}}(W)$ of the GMM estimator, we can provide asymptotically conservative estimators $\hat{V}_{\mathrm{C}}(W_N)$ and $\hat{V}_Z(W_N)$. We recommend to use the second estimator $\hat{V}_Z(W_N)$ under the finite-population asymptotic framework because it is less conservative than the first one $\hat{V}_{\mathrm{C}}(W_N)$.

2.3. **Choice of GMM weight.** The asymptotic analysis in the previous subsection focuses on the case where the weight matrix W_N for the GMM estimation is given. Given the different form of the asymptotic variance $V_{\mathrm{GMM}}(W)$ from the one under the conventional infinite-population asymptotics, it is interesting to investigate the choice of the weight matrix under the current setup.

First of all, the variance $V_{\mathrm{GMM}}(W)$ is minimized by $W_{\mathrm{opt}} = (\Omega - \rho\Delta)^{-1}$ in the matrix sense. However, due to the component Δ, a consistent estimator of W_{opt} is not available in general. Motivated by the discussion in the previous subsection, we can consider two feasible weights, $\hat{\Omega}^{-1}$ and $(\hat{\Omega} - \rho\hat{\Delta}_Z)^{-1}$. Theorems 1 and 2 imply

$$\begin{aligned} \sqrt{N}(\hat{\theta}_N(\hat{\Omega}^{-1}) - \theta_M^*) &\xrightarrow{d} N(0, V_{\mathrm{GMM}}(\Omega^{-1})), \\ \sqrt{N}(\hat{\theta}_N((\hat{\Omega} - \rho\hat{\Delta}_Z)^{-1}) - \theta_M^*) &\xrightarrow{d} N(0, V_{\mathrm{GMM}}((\Omega - \rho\Delta_Z)^{-1})). \end{aligned}$$

Although these asymptotic variances are not directly comparable, we can see that

$$V_{\mathrm{GMM}}((\Omega - \rho\Delta_Z)^{-1}) \leq_{\mathrm{pd}} V_Z((\Omega - \rho\Delta_Z)^{-1}) \leq_{\mathrm{pd}} V_Z(\Omega^{-1}) \leq_{\mathrm{pd}} V_{\mathrm{C}}(\Omega^{-1}). \tag{2.3}$$

Based on these relationships, to conduct inference on the parameters θ^*_M, we recommend to use the point estimator $\hat{\theta}_N((\hat{\Omega} - \rho\hat{\Delta}_Z)^{-1})$ combined with the asymptotic variance estimator $\hat{V}_Z((\hat{\Omega} - \rho\hat{\Delta}_Z)^{-1})$.

3. Numerical illustration

In this section, we conduct a simulation study based on real economic data. We employ the dataset studied by Lyons and Zhang (2017) and Xu (2021) on entrepreneurship training and incubation programs in North American undergraduates between 2011 and 2015. Lyons and Zhang (2017) originally analyzed the entrepreneurial activities of minority groups using data from 179 finalists out of 188 who accepted the program and 156 finalists out of 166 who did not. As in Xu (2021), we consider this data as the entire finite-population and analyze the impact of entrepreneurial activities.

The dataset contains two outcomes for entrepreneurial activities: short-term and ongoing/long-run. The short-term activity is set to one if the finalist has founded/co-founded a startup or worked for a venture capital firm or a startup after the program but is no longer working at a startup. The ongoing/longer-run startup activity takes one if the finalist is currently working with a startup. For these binary outcomes, we estimate the linear probability model (LPM) and probit model by gender and outcome types. For each analysis, we include observable fixed attributes: location, program interview scores, prior entrepreneurial experience, study major, year of study, university ranking, and year and interviewer dummies. However, as described by Lyons and Zhang (2017) and Xu (2021), unobservable attribute differences between those who accepted the program and those who did not may cause self-selection into the program. Therefore, we take the same empirical strategy and discuss mainly the magnitude of the standard errors rather than point estimates since our concern is the standard errors.

We conduct a simulation study using this data. In each Monte Carlo replication, we draw a subsample from the population by the sampling ratio $\rho = 0.8$ (in our preliminary simulations, we considered different values of ρ, but the results are similar). We compare the M-estimator with the associated standard error in Xu (2021) with our GMM estimator with the associated standard errors (i.e., $\hat{V}_{\mathrm{C}}(W_N)$ and $\hat{V}_{\mathrm{Z}}(W_N)$ in the previous section). For the GMM, as in Imbens and Lancaster (1994), we use the average outcomes by the treatment status in the population as additional moments, i.e., we calculate the average probabilities of short-term or ongoing activity for accepted and unaccepted finalists and incorporate them as additional information in the GMM estimation. We repeat this procedure 1000 times.

Table 1 reports the Monte Carlo averages and standard errors of the M-estimates $\hat{\theta}_M$ and GMM estimates $\hat{\theta}(\hat{\Omega}^{-1})$ and $\hat{\theta}((\hat{\Omega} - \rho\hat{\Delta}_Z)^{-1})$ using the weights $\hat{\Omega}^{-1}$ and $(\hat{\Omega} - \rho\hat{\Delta}_Z)^{-1}$, respectively. It reports the point estimates and the standard errors of the program treatment effect for male and female groups by using the linear probability and probit models. From the left panel of Table 1, we can see that the averages of the estimates are very close for both the M- and GMM estimates across all cases, and that program participation has a more considerable impact on females than males on both short-term and ongoing/long-term outcomes. Also the right panel of Table 1 shows that the Monte Carlo standard errors of the GMM estimators are smaller than the ones of

Table 1. Monte Carlo averages and standard errors

	Monte Carlo average			Monte Carlo standard error		
	M-estimator	GMM		M-estimator	GMM	
	$\hat{\theta}_M$	$\hat{\theta}(\hat{\Omega}^{-1})$	$\hat{\theta}((\hat{\Omega}-\rho\hat{\Delta}_Z)^{-1})$	$\hat{\theta}_M$	$\hat{\theta}(\hat{\Omega}^{-1})$	$\hat{\theta}((\hat{\Omega}-\rho\hat{\Delta}_Z)^{-1})$
Short term/LPM						
Female	0.16446	0.16402	0.16433	0.04327	0.03592	0.03625
Male	0.13854	0.13868	0.13869	0.02410	0.01363	0.01366
Ongoing/LPM						
Female	0.27722	0.27654	0.27639	0.05285	0.04382	0.04410
Male	0.09959	0.10023	0.10004	0.02889	0.01687	0.01704
Short term/Probit						
Female	1.09711	1.13677	1.14327	0.37302	0.34464	0.34966
Male	0.57978	0.56989	0.56995	0.11001	0.06281	0.06299
Ongoing/Probit						
Female	1.28117	1.28771	1.28710	0.29879	0.25940	0.26120
Male	0.28762	0.28210	0.28186	0.07961	0.04836	0.04863

the M-estimators, and that the GMM standard errors are similar for the both weights. In terms of the standard errors, efficiency gains of the GMM estimators over the M-estimator are around 6.26-17.08% for females and 38.91-43.44% for males.

It should be noted that the Monte Carlo standard errors of the M- and GMM estimators are estimating (squared root of) the asymptotic variances, which cannot be estimated consistency under the finite-population asymptotics. In Table 2 below, we present the Monte Carlo averages of the asymptotically conservative standard errors: (i) Eicker-Huber-White standard error of the M-estimator, (ii) Xu's (2021) finite-population standard error of the M-estimator, (iii) Eicker-Huber-White standard error based on $\hat{V}_{\mathrm{C}}(\hat{\Omega}^{-1})$, (iv) finite-population standard error based on $\hat{V}_{\mathrm{Z}}(\hat{\Omega}^{-1})$, and (v) finite-population standard error based on $\hat{V}_{\mathrm{Z}}((\hat{\Omega}-\rho\hat{\Delta}_Z)^{-1})$. Based on the analysis in Section 2.3, our recommendation is (v) with the GMM estimator $\hat{\theta}((\hat{\Omega}-\rho\hat{\Delta}_Z)^{-1})$.

Table 2. Monte Carlo averages of conservative standard errors

	M-estimator		GMM		
	(i)	(ii)	(iii)	(iv)	(v)
Short term/LPM					
Female	0.11259	0.09851	0.09024	0.08242	0.06483
Male	0.06657	0.06402	0.03824	0.03715	0.03431
Ongoing/LPM					
Female	0.13746	0.12020	0.11090	0.10280	0.08074
Male	0.07899	0.07604	0.04465	0.04321	0.03991
Short term/Probit					
Female	0.63175	0.56193	0.60853	0.54802	0.43037
Male	0.27185	0.26397	0.16572	0.16140	0.14906
Ongoing/Probit					
Female	0.54720	0.49208	0.51811	0.48900	0.37355
Male	0.21433	0.20790	0.12501	0.12134	0.11207

Our findings are summarized as follows. Here we focus on the estimates for females although similar comments apply to the ones for males. First, we confirm that these asymptotically conservative standard errors do not fall below the corresponding Monte Carlo standard errors in Table 1. The conventional standard errors in (i) for the M-estimators and (iii) for the GMM are around 69.36-160.20% and 76.56-153.08% larger than the (infeasible) Monte Carlo standard errors in Table 1, respectively. On the other hand, the finite-population standard errors (ii), (iv), and (v) are around 50.64-127.66%, 59.01-129.45%, and 23.08-83.08% larger than the (infeasible) Monte Carlo standard errors in Table 1, respectively. Based on these results, we can also see that the finite-population standard errors can alleviate the overestimation of the conventional standard error. For example, improvements by using the suggested finite-population standard error (v) over the conventional one (iii) are around 27.19-29.27%. Second, if we compare the finite-population standard error (ii) for the M-estimator with the recommended one (v) for the GMM estimator, we can see that (v) is smaller than (ii) by 23.41-34.18%. This result shows that using the additional moments can also help to reduce the feasible but conservative standard errors. Finally, Tables 1 and 2 indicate that significance of the coefficients may change depending on the standard errors. For example, if we look at the estimates and standard errors for females in the first rows of Tables 1 and 2, inferences based on (i)-(iii) tend to accept the null of no significance, but (iv)-(v) tend to reject at 5% significance level.

Appendix A. Mathematical appendix

A.1. **Proof of Theorem 1.**

A.1.1. *Proof of Part (1).* It is sufficient to verify the conditions in Newey and McFadden (1994, Theorem 2.1). Their condition (i) is satisfied due to uniqueness of θ^* and positive definiteness of W in Assumption (2). Their condition (ii) (i.e., compactness of the parameter space) is directly imposed.

To verify their conditions (iii) and (iv), it is sufficient to show that

$$\sup_{\theta\in\Theta}\left\|\frac{1}{N}\sum_{i=1}^{M}R_i g_i(X_i,\theta)-\lim_{M\to\infty}\frac{1}{M}\sum_{i=1}^{M}\mathbb{E}[g_i(X_i,\theta)]\right\|\xrightarrow{p}0, \tag{A.1}$$

and $\lim_{M\to\infty}\frac{1}{M}\sum_{i=1}^{M}\mathbb{E}[g_i(X_i,\theta)]$ is continuous at each $\theta\in\Theta$. The continuity of $\lim_{M\to\infty}\frac{1}{M}\sum_{i=1}^{M}\mathbb{E}[g_i(X_i,\theta)]$ follows by the dominated convergence theorem and Jensen's inequality due to the conditions on $g_i(X_i,\theta)$ in Assumption (2). For (A.1), we first note that Abadie *et al.* (2014, Lemma A.2) implies the pointwise convergence

$$\frac{1}{N}\sum_{i=1}^{M}R_i g_i(X_i,\theta)-\lim_{M\to\infty}\frac{1}{M}\sum_{i=1}^{M}\mathbb{E}[g_i(X_i,\theta)]\xrightarrow{p}0,\quad \text{for each } \theta\in\Theta.$$

Then Newey (1991, Corollary 2.2) combined with the Lipschitz condition in Assumption (2) implies the uniform convergence in (A.1).

Since we verify all the conditions of Newey and McFadden (1994, Theorem 2.1), the conclusion follows.

A.1.2. *Proof of Part (2).* Let $\hat{G}_N(\theta)=\frac{1}{N}\sum_{i=1}^{M}R_i\frac{\partial g_i(X_i,\theta)}{\partial\theta'}$. By the consistency of $\hat{\theta}_N(W_N)$ and Assumption (3) ($\theta^*\in\text{int}(\Theta)$ and differentiability of $g_i(x,\theta)$), the estimator $\hat{\theta}_N(W_N)$ satisfies the first-order condition

$$\hat{G}_N(\hat{\theta}_N(W_N))'W_N\left\{\frac{1}{N}\sum_{i=1}^{M}R_i g_i(X_i,\hat{\theta}_N(W_N))\right\}=0,$$

with probability approaching one. By expanding $g_i(X_i,\hat{\theta}_N(W_N))$ around $\hat{\theta}_N(W_N)=\theta^*_M$ and solving for $\hat{\theta}_N(W_N)-\theta^*_M$, we obtain

$$\sqrt{N}(\hat{\theta}_N(W_N)-\theta^*_M)=[\hat{G}_N(\hat{\theta}_N(W_N))'W_N\hat{G}_N(\tilde{\theta}_N)]^{-1}\hat{G}_N(\hat{\theta}_N(W_N))'W_N\frac{1}{\sqrt{N}}\sum_{i=1}^{M}R_i g_i(X_i,\theta^*_M),$$

where $\tilde{\theta}_N$ is a point on the line joining $\hat{\theta}_N(W_N)$ and θ^*_M. Since $\hat{\theta}_N(W_N)-\theta^*_M\xrightarrow{p}0$ and $\tilde{\theta}_N-\theta^*_M\xrightarrow{p}0$, it is sufficient for the conclusion to show that

$$\sup_{\theta\in\mathcal{N}}\left\|\hat{G}_N(\theta)-\mathbb{E}\left[\frac{\partial g_i(X_i,\theta)}{\partial\theta'}\right]\right\| \xrightarrow{p} 0, \tag{A.2}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{M}R_i g_i(X_i,\theta^*_M) \xrightarrow{d} N(0,\Omega-\rho\Delta), \tag{A.3}$$

for some neighborhood around θ^*.

For (A.2), Abadie *et al.* (2014, Lemma A.2) under Assumption (3) implies the pointwise convergence $\hat{G}_N(\theta) - \mathbb{E}\left[\frac{\partial g_i(X_i,\theta)}{\partial\theta'}\right] \xrightarrow{p} 0$ for each $\theta \in \mathcal{N}$. Then Newey (1991, Corollary 2.2) combined with the Lipschitz condition in Assumption (3) implies the uniform convergence in (A.2). For (A.3), it follows directly from Abadie *et al.* (2020, Lemma A.1) under Assumptions (2)-(3). Therefore, the conclusion follows.

References

[1] Abadie, A., Athey, S., Imbens, G. W. and J. M. Wooldridge (2014) Finite population causal standard errors, Working paper no. 20325, National Bureau of Economic Research.

[2] Abadie, A., Athey, S., Imbens, G. W. and J. M. Wooldridge (2020) Sampling-based versus design-based uncertainty in regression analysis, *Econometrica*, 88, 265-296.

[3] Athey, S. and G. W. Imbens, (2022) Design-based analysis in difference-in-differences settings with staggered adoption, *Journal of Econometrics*, 226, 62-79.

[4] Duflo, E. and A. Banerjee (2017) *Handbook of field experiments*, Elsevier.

[5] Freedman, D. A. (2008a) On regression adjustments in experiments with several treatments, *Annals of Applied Statistics*, 2, 176-96.

[6] Freedman, D. A. (2008b) On regression adjustments to experimental data, *Advances in Applied Mathematics,* 40, 180-93.

[7] Hall, A. R. (2005) *Generalized Method of Moments*, Oxford University Press.

[8] Imbens, G. W. and T. Lancaster (1994) Combining micro and macro data in microeconometric models, *Review of Economic Studies*, 61, 655-680.

[9] Lyons, E. and L. Zhang (2017) The impact of entrepreneurship programs on minorities, *American Economic Review: Papers and Proceedings*, 107, 303-7.

[10] Muralidharan, K. and P. Niehaus (2017) Experimentation at scale, *Journal of Economic Perspectives*, 31, 103-124.

[11] Newey, W. K. (1991) Uniform convergence in probability and stochastic equicontinuity, *Econometrica*, 59, 1161-7.

[12] Newey, W. K. and D. L. McFadden (1994) Large sample estimation and hypothesis testing, in Engle, R. F. and D. L. McFadden (eds.), *Handbook of Econometrics*, vol. 4, 2111-2245, Elsevier.

[13] Neyman, J. (1923) On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Annals of Agricultural Sciences,* 10, 1–5

[14] Prášková, Z. and P. K. Sen (2009) Asymptotics in finite population sampling, *Handbook of Statistics*, 29, 489-522, Elsevier.

[15] Rambachan, A. and J. Roth (2020) Design-based uncertainty for quasi-experiments, *arXiv preprint arXiv:2008.00602.*

[16] Rosenbaum, P. R. (2002) *Observational studies*, Springer.

[17] Roth, J. and P. H. Sant'Anna (2023) Efficient estimation for staggered rollout designs, forthcoming in *Journal of Political Economy.*

[18] Xu, R. (2021) Potential outcomes and finite-population inference for M-estimators, *Econometrics Journal*, 24, 162-176.

[19] Xu, R. and J. M. Wooldridge (2022) A design-based approach to spatial correlation, *arXiv preprint arXiv:2211.14354.*

Graduate School of Economics, Keio University, 2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan.

Email address: haruokakehi@keio.jp

Department of Economics, London School of Economics, Houghton Street, London, WC2A 2AE, UK, and Keio Economic Observatory (KEO), 2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan.

Email address: t.otsu@lse.ac.uk