

CREA Discussion Paper 2017-20

Economics

Center for Research in Economics and Management
University of Luxembourg

Inference in Conditional Moment Restriction Models when there is Selection due to Stratification

available online : http://www.wfr.uni.lu/recherche/fdef/crea/publications2/discussion_papers

Antonio Cosma, CREA, Université du Luxembourg
Andrei Kostyrka, CREA, Université du Luxembourg
Gautham Tripathi, CREA, Université du Luxembourg

October, 2017

For editorial correspondence, please contact: crea@uni.lu
University of Luxembourg
Faculty of Law, Economics and Finance
162A, avenue de la Faiencerie
L-1511 Luxembourg

INFERENCE IN CONDITIONAL MOMENT RESTRICTION MODELS WHEN THERE IS SELECTION DUE TO STRATIFICATION

ANTONIO COSMA, ANDREĪ KOSTYRKA, AND GAUTAM TRIPATHI

ABSTRACT. We show how to conduct efficient semiparametric inference in models specified as conditional moment equalities when data is collected by variable probability sampling.

1. INTRODUCTION

The golden standard for collecting data, at least for the ease of doing subsequent statistical analysis, is simple random sampling, whereby each observation in the “target” population, namely, the population of interest, has an equal chance of being chosen. Consequently, the probability distribution of the chosen observation, regarded as belonging to a “realized” population, is the same as the probability distribution of an observation in the target population, which facilitates statistical analysis.

However, when estimating or testing economic relationships, economists often discover that the data they plan to use is not drawn from the target population they wish to study. Instead, the observations are found to be sampled from a related but different distribution. Sometimes this is done deliberately to make the sample more informative. E.g., when studying the impact of welfare legislation, it is desirable to oversample minorities and low income families. Similarly, if we want to examine the effect of disability laws on demand for public transportation, it makes sense to oversample households with handicapped members. At other times, a distinction between the target and realized populations can be created unintentionally. E.g., in sampling the duration of unemployment at a randomly chosen time, economists are more likely to observe longer unemployment spells than shorter ones. As alluded to in Newey (1993, p. 419), using a dataset to answer questions for which it was not originally designed — a typical situation in economics where data is costly to collect — may also lead to such a situation. For instance, if the reason for collecting data is to estimate mean income for an underlying population, oversampling low income and undersampling high income families can improve the precision of estimators. However, at some later stage this income data can be used by another researcher as the dependent variable in a regression model without realizing that the original sample was drawn from a distribution other than the target population.

Whatever its cause, if the distinction between the target and realized populations is not taken into account when analyzing the data, statistical inference can be seriously off the mark. This phenomenon is commonly called selection bias. Cf. Heckman (1976, 1979) and Manski (1989, 1995) for a classic exposition of the selection problem.

In this paper, we describe an efficient semiparametric approach for conducting inference in conditional moment restriction models when data is collected by a variable probability sampling scheme such that the observations from the target population have unequal chances of being chosen. In other words, we show how to efficiently deal with the selection bias caused by the sampling scheme used to collect the data, because the sampling scheme induces a probability distribution on the realized population which differs from the target distribution for which inference is to be made.

2. THE MODEL

2.1. Conditional moment equalities. Let $Z^* := (Y^*, X^*)_{\dim(Y^*) + \dim(X^*) \times 1}$ be a random (column) vector that denotes an observation from the target population, where Y^* is the vector of endogenous variables and X^* the vector of exogenous variables. Assume that

$$H_0 : \quad \exists \theta^* \in \mathbb{R}^{\dim(\theta^*)} \quad \text{s.t.} \quad \mathbb{E}_{P_{Y^*|X^*}^*} [g(Z^*, \theta^*) | X^*] = 0 \quad P_{X^*}^* \text{-a.s.}, \quad (2.1)$$

where g is a vector of functions known up to θ^* , the notation $\mathbb{E}_{P_{Y^*|X^*}^*}$ indicates that the conditional expectation is with respect to the conditional distribution $P_{Y^*|X^*}^* := \text{Law}(Y^* | X^*)$, and $P_{X^*}^*$ denotes the marginal distribution of X^* . The conditional distribution of $Y^* | X^*$ and the marginal distribution of X^* are unknown.¹ Throughout the paper, random variables and probability measures associated with the target population appear with the superscript “*.” The parameter of interest θ^* has an asterisk attached to it because it is a functional of $P_{Y^*|X^*}^*$.²

A large class of models in applied economics can be characterized in terms of conditional moment equalities of the form (2.1). E.g., in linear regression models where some of the regressors are endogenous, we have $g(Z^*, \theta^*) := Y_1^* - \alpha^* - X^{*'}\beta^* - Y_2^{*'}\delta^*$, where $Y^* = (Y_1^*, Y_2^*)$, Y_1^* is the outcome variable, Y_2^* the vector of endogenous regressors, X^* the vector of exogenous regressors, and $\theta^* = (\alpha^*, \beta^*, \delta^*)$. Similarly, for nonlinear regression models, $g(Z^*, \theta^*) := Y_1^* - \psi(Y_2^*, X^*, \theta^*)$, where the nonlinear function $\psi(Y_2^*, X^*, \cdot)$ is known up to θ^* . Multivariate extensions include systems of equations or transformation models, linear or nonlinear, of the form $g(Z^*, \theta^*) = \varepsilon^*$, where g is a vector of known functions and the identifying assumption is that $\mathbb{E}_{P_{Y^*|X^*}^*} [\varepsilon^* | X^*] = 0$ $P_{X^*}^*$ -a.s.. Several examples of econometric models

¹If X^* is constant $P_{X^*}^*$ -a.s., then there is no conditioning and (2.1) reduces to a system of unconditional moment equalities. These models are studied in Tripathi (2011a,b).

²Similar notation, but without the “*” superscript, applies to the random variables and probability measures in the realized population.

defined via conditional moment restrictions may be found in Newey (1993, Section 3), Pagan and Ullah (1999, Chapter 3), and Wooldridge (2010).

2.2. Variable probability sampling. Instead of observing Z^* directly from the target population, we possess a random vector $Z := (Y, X)$ that is collected by variable probability (VP) sampling, also known as Bernoulli sampling. For more on VP and other stratified sampling schemes, cf., e.g., DeMets and Halperin (1977), Manski and Lerman (1977), Holt, Smith, and Winter (1980), Cosslett (1981a,b, 1991, 1993), Manski and McFadden (1981), Jewell (1985), Quesenberry and Jewell (1986), Scott and Wild (1986), Kalbfleisch and Lawless (1988), Bickel and Ritov (1991), Imbens (1992), Imbens and Lancaster (1996), Deaton (1997), Wooldridge (1999, 2001), Butler (2000), Bhattacharya (2005, 2007), Hirose (2007), Hirose and Lee (2008), Tripathi (2011a,b), and Severini and Tripathi (2013).

Let the support of Z^* , denoted by $\text{supp}(Z^*)$, be partitioned into L nonempty disjoint strata $\mathbb{C}_1, \dots, \mathbb{C}_L$. In VP sampling, typically used when data is collected by telephone surveys, an observation is first drawn randomly from the target population. If it lies in stratum \mathbb{C}_l , it is retained with known probability $p_l \in (0, 1]$. If it is discarded, all information about the observation is lost. Hence, instead of observing a random vector Z^* drawn from the target distribution $P^* := \text{Law}(Z^*)$, we observe a random vector Z , with $\text{supp}(Z) = \text{supp}(Z^*)$, drawn from the realized distribution $P := \text{Law}(Z)$ given by³

$$P(Z \in B) := \sum_{l=1}^L \frac{p_l}{b^*} \int_B \mathbb{1}_{\mathbb{C}_l}(z) dP^*(z), \quad B \in \mathcal{B}(\mathbb{R}^{\dim(Z^*)}), \quad (2.2)$$

where $\mathcal{B}(\mathbb{R}^{\dim(Z^*)})$ is the Borel sigma-field of $\mathbb{R}^{\dim(Z^*)}$, $b^* := \sum_{l=1}^L p_l Q_l^*$, and $Q_l^* := P^*(Z^* \in \mathbb{C}_l) > 0$ denotes the probability that a randomly chosen observation from the target population lies in the l th stratum.

Since Q_l^* represents the probability mass of the l th stratum in the target population, the Q_l^* 's are popularly called “aggregate shares.” The aggregate shares, which add up to one, i.e., $\sum_{l=1}^L Q_l^* = 1$, are unknown parameters of interest to be estimated along with the structural parameter θ^* . The parameter b^* also has a practical interpretation, namely, it is the probability that an observation drawn from the target population during the sampling process is ultimately retained in the sample.

³Cf. Severini and Tripathi (2013, Appendix H) for a short proof of (2.2).

It is immediate from (2.2) that the density of P , with respect to any measure on $\mathcal{B}(\mathbb{R}^{\dim(Z^*)})$ that dominates P^* , is given by

$$\begin{aligned} dP(z) &:= \sum_{l=1}^L \frac{p_l}{b^*} \mathbb{1}_{C_l}(z) dP^*(z) & (z \in \mathbb{R}^{\dim(Z^*)}) \\ &= \frac{b(z)}{b^*} dP^*(z), & (2.3) \end{aligned}$$

where $b(z) := \sum_{l=1}^L p_l \mathbb{1}_{C_l}(z)$. Following Imbens and Lancaster (1996, p. 296), $b(\cdot)/b^*$ is referred to as a bias function because it determines the selection bias due to stratified sampling, i.e., the extent to which P differs from P^* . For instance, it is easy to see that if the sampling probabilities p_1, \dots, p_L are all equal, then there is no selection bias, i.e., $P = P^*$, because $b(\cdot)/b^* = 1$ irrespective of the values taken by the aggregate shares.

The marginal density of X is given by

$$\begin{aligned} dP_X(x) &:= \int_{y \in \mathbb{R}^{\dim(Y^*)}} dP(y, x) & (x \in \mathbb{R}^{\dim(X^*)}) \\ &= \int_{y \in \mathbb{R}^{\dim(Y^*)}} \frac{b(y, x)}{b^*} dP_{Y^*|X^*=x}^*(y) dP_{X^*}^*(x) & ((2.3)) \\ &= \frac{\gamma^*(x)}{b^*} dP_{X^*}^*(x), & (2.4) \end{aligned}$$

where $\gamma^*(x) := \mathbb{E}_{P_{Y^*|X^*}^*}[b(Y^*, x)|X^* = x]$. Throughout the paper, we maintain the assumption that $\gamma^* > 0$ on $\text{supp}(X^*)$.⁴ Under this condition, the probability distributions P_X and $P_{X^*}^*$ are mutually absolutely continuous, which we denote by writing $P_{X^*}^* \ll P_X \ll P_{X^*}^*$.

Since $\text{supp}(Y, X) = \text{supp}(Y^*, X^*)$ and $\gamma^* > 0$ on $\text{supp}(X^*)$, the conditional density of $Y|X$ is given by

$$\begin{aligned} dP_{Y|X=x}(y) &:= \frac{dP(y, x)}{dP_X(x)} & ((y, x) \in \text{supp}(Y^*) \times \text{supp}(X^*)) \\ &= \frac{b(y, x)}{\gamma^*(x)} dP_{Y^*|X^*=x}^*(y), & (2.5) \end{aligned}$$

where (2.5) follows from (2.3) and (2.4).

By (2.5), $dP_{Y|X=x}(y) = dP_{Y^*|X^*=x}^*(y)$ if and only if $b(y, x) = \gamma^*(x)$ for all $(y, x) \in \text{supp}(Y^*) \times \text{supp}(X^*)$. However, as discussed subsequently, the condition $b(y, x) = \gamma^*(x)$ holds only in a special case. Therefore, in general, $dP_{Y|X} \neq dP_{Y^*|X^*}^*$. Consequently, estimating (2.1) using the realized sample without accounting for the fact that it was obtained by stratified sampling, i.e., ignoring stratification, will generally not lead to a consistent estimator of θ^* .

⁴A sufficient condition for this is that $P_{Y^*|X^*}^*((Y^*, x) \in C_l|X^* = x) > 0$ for each l and $x \in \text{supp}(X^*)$.

2.3. Endogenous and exogenous stratification. As noted by Wooldridge (1999, p. 1385), VP sampling is employed when it is cheaper to obtain information on a subset of variables in the target population. Hence, it may happen that in certain datasets only Y^* is stratified (endogenous stratification),⁵ or only X^* is stratified (exogenous stratification), or both Y^* and X^* are stratified. To see how all these cases can be handled in a unified manner in our framework, let the support of Y^* be partitioned into J nonempty disjoint strata $\mathbb{A}_1, \dots, \mathbb{A}_J$, and the support of X^* be partitioned into M nonempty disjoint strata $\mathbb{B}_1, \dots, \mathbb{B}_M$. Then, since $\cup_{j=1}^J \mathbb{A}_j \times \cup_{m=1}^M \mathbb{B}_m = \cup_{j=1}^J \cup_{m=1}^M \mathbb{A}_j \times \mathbb{B}_m$,

$$\text{supp}(Y^*, X^*) = \begin{cases} \cup_{j=1}^J \cup_{m=1}^M \mathbb{A}_j \times \mathbb{B}_m & \text{if both } Y^* \text{ and } X^* \text{ are stratified} \\ \cup_{j=1}^J (\mathbb{A}_j \times \text{supp}(X^*)) & \text{if only } Y^* \text{ is stratified} \\ \cup_{m=1}^M \text{supp}(Y^*) \times \mathbb{B}_m & \text{if only } X^* \text{ is stratified.} \end{cases}$$

Therefore, if both Y^* and X^* are stratified, then $\text{supp}(Z^*) = \cup_{l=1}^L \mathbb{C}_l$ with $L = JM$ and each $\mathbb{C}_l = \mathbb{A}_j \times \mathbb{B}_m$ for some $(j, m) \in \{1, \dots, J\} \times \{1, \dots, M\}$. This is the most general case for which $P_{Y|X}$ is given by (2.5).⁶

In contrast, simplifications occur if the stratification is endogenous or exogenous. For instance, if stratification is based only on Y^* , then $\text{supp}(Z^*) = \cup_{l=1}^L \mathbb{C}_l$ with $L = J$ and $\mathbb{C}_l = \mathbb{A}_l \times \text{supp}(X^*)$, which implies that $b(y, x) = \sum_{l=1}^J p_l \mathbb{1}_{\mathbb{A}_l \times \text{supp}(X^*)}(y, x)$. Hence, by (2.5),

$$\text{endogenous stratification} \implies \begin{cases} dP_{Y|X=x}(y) = \frac{b(y)}{\gamma^*(x)} dP_{Y^*|X^*=x}(y) \\ b(y) := \sum_{l=1}^J p_l \mathbb{1}_{\mathbb{A}_l}(y) \\ \gamma^*(x) = \mathbb{E}_{P_{Y^*|X^*}}[b(Y^*)|X^* = x] \end{cases} \quad (2.6)$$

for $(y, x) \in \text{supp}(Y^*) \times \text{supp}(X^*)$.

If stratification is based only on X^* , then $\text{supp}(Z^*) = \cup_{l=1}^L \mathbb{C}_l$ with $L = M$ and $\mathbb{C}_l = \text{supp}(Y^*) \times \mathbb{B}_l$, which implies that $b(y, x) = \sum_{l=1}^M p_l \mathbb{1}_{\text{supp}(Y^*) \times \mathbb{B}_l}(y, x)$. Hence, by (2.5),

$$\text{exogenous stratification} \implies \begin{cases} dP_{Y|X=x}(y) = dP_{Y^*|X^*=x}(y) \\ b(x) := \sum_{l=1}^M p_l \mathbb{1}_{\mathbb{B}_l}(x) \\ \gamma^*(x) = b(x) \end{cases} \quad (2.7)$$

for $(y, x) \in \text{supp}(Y^*) \times \text{supp}(X^*)$.

Consequently, exogenous stratification can be ignored, at least as far as consistent estimation is concerned. However, as the following example demonstrates, ignoring endogenous stratification does not lead to a consistent estimator.

⁵In the econometrics literature, stratification based on a finite set of response variables is often referred to as choice based sampling.

⁶Unless mentioned otherwise, it is assumed throughout the paper that both Y^* and X^* are stratified.

Example 2.1 (Linear regression with exogenous regressors). Consider the linear regression model $Y^* = \widetilde{X}^* \theta^* + \varepsilon^*$, where $\widetilde{X}^* := (1, X^*)$. Assume that the regressors are exogenous with respect to the model error in the target population, i.e., $\mathbb{E}_{P_{Y^*|X^*}}[\varepsilon^*|X^*] = 0$ P_{X^*} -a.s..

Suppose that only Y^* is stratified. If we ignore the fact that the data were collected by VP sampling and simply regress the observed Y on the observed X , then θ^* cannot be consistently estimated by the least-squares (LS) estimator. To see this, let $\widetilde{X} := (1, X)$ and observe that

$$\begin{aligned} \mathbb{E}_{P_{Y|X}}[Y|X = x] &\stackrel{(2.6)}{=} \frac{1}{\gamma^*(x)} \mathbb{E}_{P_{Y^*|X^*}}[Y^* b(Y^*)|X^* = x] && (x \in \text{supp}(X^*)) \\ &= \frac{1}{\gamma^*(x)} \mathbb{E}_{P_{Y^*|X^*}}[(\widetilde{X}^* \theta^* + \varepsilon^*) b(Y^*)|X^* = x] \\ &= \widetilde{x}' \theta^* + \frac{1}{\gamma^*(x)} \mathbb{E}_{P_{Y^*|X^*}}[\varepsilon^* b(Y^*)|X^* = x] && (2.8) \\ &\neq \widetilde{x}' \theta^* \quad \text{with positive } P_{X^*}\text{-probability,} \end{aligned}$$

because $\mathbb{E}_{P_{Y^*|X^*}}[\varepsilon^* b(Y^*)|X^*] \neq 0$ with positive P_{X^*} -probability. Hence, regressing Y on \widetilde{X} will not yield a consistent estimator of θ^* . Indeed, letting $\hat{\theta}_{\text{LS}}$ denote the LS estimator obtained by regressing Y on \widetilde{X} ,

$$\begin{aligned} \text{plim}_{n \rightarrow \infty} \hat{\theta}_{\text{LS}} &= \text{plim}_{n \rightarrow \infty} (n^{-1} \sum_{j=1}^n \widetilde{X}_j \widetilde{X}_j')^{-1} (n^{-1} \sum_{j=1}^n \widetilde{X}_j Y_j) \\ &= (\mathbb{E}_{P_X} \widetilde{X} \widetilde{X}')^{-1} (\mathbb{E}_P \widetilde{X} Y) \\ &= (\mathbb{E}_{P_X} \widetilde{X} \widetilde{X}')^{-1} (\mathbb{E}_{P_X} \widetilde{X} \mathbb{E}_{P_{Y|X}}[Y|X]) && (2.9) \\ &= (\mathbb{E}_{P_X} \frac{\gamma^*(X^*)}{b^*} \widetilde{X}^* \widetilde{X}^{*'})^{-1} (\mathbb{E}_{P_X} \frac{\gamma^*(X^*)}{b^*} \widetilde{X}^* [\widetilde{X}^* \theta^* + \frac{1}{\gamma^*(X^*)} \mathbb{E}_{P_{Y^*|X^*}}[\varepsilon^* b(Y^*)|X^*]]) \\ &= \theta^* + (\mathbb{E}_{P_X} \gamma^*(X^*) \widetilde{X}^* \widetilde{X}^{*'})^{-1} (\mathbb{E}_{P_X} \widetilde{X}^* \varepsilon^* b(Y^*)) \\ &\neq \theta^*, \end{aligned}$$

where the fourth equality follows from (2.4) and (2.8).

Of course, if stratification is exogenous, then

$$\begin{aligned} \mathbb{E}_{P_{Y|X}}[Y|X = x] &\stackrel{(2.7)}{=} \mathbb{E}_{P_{Y^*|X^*}}[Y^*|X^* = x] && (x \in \text{supp}(X^*)) \\ &= \mathbb{E}_{P_{Y^*|X^*}}[\widetilde{X}^* \theta^* + \varepsilon^*|X^* = x] \\ &= \widetilde{x}' \theta^* \quad \text{for } P_{X^*}\text{-a.a. } x \in \text{supp}(X^*). && (2.10) \end{aligned}$$

Hence, ignoring exogenous stratification does not affect the consistency of $\hat{\theta}_{\text{LS}}$ because

$$\begin{aligned} \text{plim}_{n \rightarrow \infty} \hat{\theta}_{\text{LS}} &= (\mathbb{E}_{P_X} \tilde{X} \tilde{X}')^{-1} (\mathbb{E}_{P_X} [\tilde{X} \mathbb{E}_{P_{Y|X}}(Y|X)]) && ((2.9)) \\ &= (\mathbb{E}_{P_X^*} \frac{\gamma^*(X^*)}{b^*} \tilde{X}^* \tilde{X}^{*'})^{-1} (\mathbb{E}_{P_X^*} \frac{\gamma^*(X^*)}{b^*} \tilde{X}^* \tilde{X}^{*'}) \theta^* && ((2.4) \ \& \ (2.10)) \\ &= \theta^*. \end{aligned}$$

However, as shown subsequently (cf. Example 3.1), $\hat{\theta}_{\text{LS}}$ is not asymptotically efficient. Hence, ignoring exogenous stratification does not affect the consistency of the LS estimator,⁷ but it does affect its efficiency. \square

3. INFERENCE

3.1. Related literature and our contribution. There is a large literature on estimation and testing models using data collected by various types of stratified sampling schemes; cf. the papers cited at the beginning of Section 2.2, and the references therein. In this section, we briefly describe only some of the works that consider VP sampling.

Earlier papers in the literature on estimating models with conditioning variables assume that $P_{Y^*|X^*}^*$ is known up to a finite dimensional parameter; only $P_{X^*}^*$ is left completely unspecified. E.g., a well known application of VP sampling can be found in Hausman and Wise (1981). Imbens and Lancaster (1996) extend the maximum likelihood approach of Hausman and Wise to a moment based methodology that allows for VP sampling, mixed response variables, and stratification on exogenous covariates. Regression under VP sampling and a parametric $P_{Y^*|X^*}^*$ has also been investigated. E.g., Jewell (1985) and Quesenberry and Jewell (1986) propose iterative estimators of regression coefficients under VP sampling without imposing normality or independence, though they do not provide any asymptotic theory for their estimators.

The papers described above impose strong conditions on the distribution of $Y^*|X^*$. Exceptions include Wooldridge (1999) and Tripathi (2011b), who leave both $P_{Y^*|X^*}^*$ and $P_{X^*}^*$ completely unspecified. Wooldridge provides asymptotic theory for M -estimation under VP sampling for a model defined in terms of a set of just-identified unconditional moment equalities, whereas Tripathi considers optimal generalized method of moments (GMM) estimation in unconditional moment restriction models that allow for the parameter of interest to be over-identified. The major difference between (2.1) and the models in the papers of Wooldridge and Tripathi is that (2.1) is a conditional moment restriction, whereas the moment conditions in the aforementioned papers are all unconditional. Therefore, (2.1) nests the moment conditions of Wooldridge and Tripathi as a special case.

⁷Tripathi (2011b) shows that in unconditional moment restriction models even exogenous stratification cannot be ignored.

In this paper, we show how to efficiently estimate θ^* and the aggregate shares using a smoothed empirical likelihood based approach. The results presented here answer the question posed in Wooldridge (1999, p. 1402) by providing efficiency bounds for models with conditional moment restrictions under VP sampling and showing that these bounds are attainable.

Furthermore, the results in this paper are also directly applicable to a class of “biased sampling” problems. To see this, recall that the phenomenon where the realized probability distribution P differs from the target probability distribution P^* is generically referred to as selection bias.⁸ It is useful to note that the class of problems that can be handled when selection is modeled using (2.3) includes more than just those involving stratified sampling. For instance, consider the so called “length biased sampling” problem where the probability of observing a random variable is proportional to its “size.” E.g., economists are more likely to observe longer unemployment spells if they are sampled at a randomly chosen time. Similarly, as Owen (2001, p. 127) points out, if internet log files are sampled randomly then longer sessions are likely to be over-represented. It is useful to examine length biased sampling in the context of VP sampling, because in length biased sampling we have

$$dP(z) := \frac{\|z\|}{\mathbb{E}_{P^*}\|Z^*\|} dP^*(z), \quad z \in \mathbb{R}^{\dim(Z^*)},$$

where $\|\cdot\|$ is the Euclidean norm. That is, length biased sampling can be expressed as (2.3) with $b(z) := \|z\|$ and $b^* := \mathbb{E}_{P^*}\|Z^*\|$. Therefore, with only minor notational changes, the results obtained in this paper can be extended to length biased sampling as well.

Length biased sampling has been extensively studied for the parametric case, i.e., where dP^* is specified up to a finite dimensional parameter. Cf., e.g., Patil and Rao (1977, 1978), Bickel, Klassen, Ritov, and Wellner (1993, Section 4.4), and Owen (2001, Chapter 6). As far as a nonparametric treatment of length biased sampling is concerned, Vardi (1982) deals with the case when P^* is unknown. Vardi assumes that both P^* and P can be sampled with positive probability. Using two independent samples (one each from P^* and P), he shows how to construct the nonparametric maximum likelihood estimators (NPMLE) of P^* and P , and also obtains their asymptotic distributions. Vardi (1985), and Gill, Vardi, and Wellner (1988), provide conditions for the existence and uniqueness of the NPMLE of P^* in a general setup when more than two independent samples from F^* and F are available. These papers concentrate on the distributions P^* and P ; there are no other parameters to estimate. Qin (1993) uses the empirical likelihood approach to construct a nonparametric likelihood ratio confidence interval for $\theta^* := \mathbb{E}_{P^*}Z^*$, i.e., a just-identified unconditional moment equality, using an independent sample from P^* and P . El-Barmi and Rothmann (1998) generalize Qin’s treatment to handle models with overidentified unconditional moment restrictions of the form $\mathbb{E}_{P^*}g(Z, \theta^*) = 0$.

⁸Hence, for the LS estimator in Example 2.1, one can say that it is inconsistent because of selection bias due to endogenous stratification, whereas exogenous stratification does not lead to any selection bias.

They also obtain efficient estimators of P^* and P . However, they do not consider the testing of overidentifying restrictions.

3.2. Efficiency bounds. The efficiency bounds for estimating θ^* and related functionals have been derived in Severini and Tripathi (2013, Section 14.3). In this section, we describe some of these bounds and discuss their salient features. Construction of estimators that achieve these bounds is considered in the next section.

As noted after (2.4), the assumption that $\gamma^* > 0$ on $\text{supp}(X^*)$ implies that the distributions P_X and $P_{X^*}^*$ are mutually absolutely continuous. Hence,

$$\begin{aligned}
(2.1) &\iff \mathbb{E}_{P_{Y^*|X^*}^*}[g(Y^*, x, \theta^*)|X^* = x] = 0 \quad \text{for } P_{X^*}^*\text{-a.a. } x \in \text{supp}(X^*) \\
&\iff \gamma^*(x)\mathbb{E}_{P_{Y|X}}\left[\frac{g(Y, x, \theta^*)}{b(Y, x)}|X = x\right] = 0 \quad \text{for } P_{X^*}^*\text{-a.a. } x \in \text{supp}(X^*) \quad ((2.5)) \\
&\iff P_{X^*}^*\{x \in \text{supp}(X^*) : \mathbb{E}_{P_{Y|X}}\left[\frac{g(Y, x, \theta^*)}{b(Y, x)}|X = x\right] \neq 0\} = 0 \quad (\gamma^* > 0) \\
&\iff P_X\{x \in \text{supp}(X^*) : \mathbb{E}_{P_{Y|X}}\left[\frac{g(Y, x, \theta^*)}{b(Y, x)}|X = x\right] \neq 0\} = 0 \quad (P_{X^*}^* \ll P_X \ll P_{X^*}^*) \\
&\iff \mathbb{E}_{P_{Y|X}}\left[\frac{g(Y, x, \theta^*)}{b(Y, x)}|X = x\right] = 0 \quad \text{for } P_X\text{-a.a. } x \in \text{supp}(X^*).
\end{aligned}$$

Therefore, letting $\rho_1(Z, \theta) := g(Z, \theta)/b(Z)$, we have that

$$(2.1) \iff \mathbb{E}_{P_{Y|X}}[\rho_1(Z, \theta^*)|X] = 0 \quad P_X\text{-a.s.} \quad (3.1)$$

Since the right hand side of (3.1) is a conditional moment equality with respect to the realized conditional distribution $P_{Y|X}$, the efficiency bound for θ^* follows from Chamberlain (1987). Namely, the efficiency bound for estimating θ^* is given by⁹

$$\text{l.b.}(\theta^*) := (\mathbb{E}_{P_X} D'(X) V_1^{-1}(X) D(X))^{-1}, \quad (3.2)$$

where $D(X) := \partial_\theta \mathbb{E}_{P_{Y|X}}[\rho_1(Z, \theta^*)|X]$ and $V_1(X) := \mathbb{E}_{P_{Y|X}}[\rho_1(Z, \theta^*) \rho_1'(Z, \theta^*)|X]$.

The right hand side of (3.2) can be expressed in terms of P^* , which helps provide additional insight into the bound. In particular, from (2.5) it is straightforward to see that

$$\begin{aligned}
D(x) &= \frac{1}{\gamma^*(x)} \partial_\theta \mathbb{E}_{P_{Y^*|X^*}^*}[g(Z^*, \theta^*)|X^* = x], \quad x \in \text{supp}(X^*), \\
V_1(x) &= \frac{1}{\gamma^*(x)} \mathbb{E}_{P_{Y^*|X^*}^*}\left[\frac{g(Z^*, \theta^*) g'(Z^*, \theta^*)}{b(Y^*, x)}|X^* = x\right].
\end{aligned} \quad (3.3)$$

Hence, by (2.4) and (3.3), the efficiency bound for estimating θ^* can also be written as

$$\text{l.b.}(\theta^*) = b^*(\mathbb{E}_{P_{X^*}^*}(\partial_\theta \mathbb{E}_{P_{Y^*|X^*}^*}[g(Z^*, \theta^*)|X^*])' V_b^{*-1}(X^*) (\partial_\theta \mathbb{E}_{P_{Y^*|X^*}^*}[g(Z^*, \theta^*)|X^*]))^{-1}, \quad (3.4)$$

⁹The abbreviation ‘‘l.b.’’ stands for ‘‘lower bound,’’ because the efficiency bound is the greatest lower bound for the asymptotic variance of any $n^{1/2}$ -consistent regular estimator.

where $V_b^*(X^*) := \mathbb{E}_{P_{Y^*|X^*}}[g(Z^*, \theta^*)g'(Z^*, \theta^*)/b(Z^*)|X^*]$.

The efficiency bound when only Y^* is stratified follows from (3.4) on replacing $b(Z^*)$ in the definition of $V_b^*(X^*)$ with $b(Y^*)$, where $b(Y^*)$ is defined in (2.6). Similarly, the bound when only X^* is stratified follows from (3.4) on replacing $b(Z^*)$ in $V_b^*(X^*)$ with $b(X^*)$, where the latter is defined in (2.7).

If there is no conditioning in (2.1), i.e., X^* is constant P_{X^*} -a.s., and $\dim(g) \geq \dim(\theta^*)$, then (3.4) reduces to the efficiency bound for estimating θ^* in unconditional moment restriction models when observations are collected by VP sampling (Severini and Tripathi, 2013, Section 14.2.1).

At the other extreme, if there is no stratification, i.e., $L = 1 = p_1$ and $\mathbb{C}_1 = \text{supp}(Z^*)$, so that $Z^* = Z$ and $P^* = P$, then the efficiency bound in (3.4) becomes

$$(\mathbb{E}(\partial_\theta \mathbb{E}[g(Z^*, \theta^*)|X^*])'(\mathbb{E}[g(Z^*, \theta^*)g'(Z^*, \theta^*)|X^*])^{-1}(\partial_\theta \mathbb{E}[g(Z^*, \theta^*)|X^*]))^{-1},$$

which is Chamberlain's 1987 bound for estimating θ^* in the absence of any selection.

Example 3.1 (Example 2.1 contd.). Here, $g(Z^*, \theta) = Y^* - \widetilde{X}^{*'}\theta$ and the efficiency bound for estimating θ^* is given by

$$\text{l.b.}(\theta^* \text{ in Example 2.1}) \stackrel{(3.4)}{=} b^* \left(\mathbb{E}_{P_{X^*}} \frac{\widetilde{X}^* \widetilde{X}^{*'}}{V_b^*(X^*)} \right)^{-1} \stackrel{(2.4),(2.5)}{=} \left(\mathbb{E}_{P_X} \frac{\widetilde{X} \widetilde{X}'}{\gamma^{*2}(X) V_1(X)} \right)^{-1}. \quad (3.5)$$

If stratification is endogenous, then

$$\text{l.b.}(\theta^* \text{ in Example 2.1})|_{\text{endog. strat.}} \stackrel{(3.5)}{=} \left(\mathbb{E}_{P_X} \frac{\widetilde{X} \widetilde{X}'}{\gamma^{*2}(X) V_{1,\text{endog}}(X)} \right)^{-1},$$

where $V_{1,\text{endog}}(X) := \mathbb{E}_{P_{Y|X}}[(Y - \widetilde{X}'\theta^*)^2/b^2(Y)|X]$.

In contrast, if stratification is exogenous then

$$\text{l.b.}(\theta^* \text{ in Example 2.1})|_{\text{exog. strat.}} \stackrel{(3.5)}{=} \left(\mathbb{E}_{P_X} \frac{\widetilde{X} \widetilde{X}'}{V_{1,\text{exog}}(X)} \right)^{-1}, \quad (3.6)$$

where $V_{1,\text{exog}}(X) := \mathbb{E}_{P_{Y|X}}[(Y - \widetilde{X}'\theta^*)^2|X]$.

Recall from Example 2.1 that, under exogenous stratification, the LS estimator consistently estimates θ^* . Since $n^{1/2}(\hat{\theta}_{\text{LS}} - \theta^*)$ is asymptotically (as $n \rightarrow \infty$) normal with mean zero and variance $V_{\text{LS,exog}} := (\mathbb{E}_{P_X} \widetilde{X} \widetilde{X}')^{-1}(\mathbb{E}_{P_X} \widetilde{X} \widetilde{X}' V_{1,\text{exog}}(X))(\mathbb{E}_{P_X} \widetilde{X} \widetilde{X}')^{-1}$, an application of a matrix version of the Cauchy-Schwarz inequality reveals that

$$\text{l.b.}(\theta^* \text{ in Example 2.1})|_{\text{exog. strat.}} \leq_L V_{\text{LS,exog}},$$

where \leq_L is the usual (Löwner) order on the set of symmetric matrices.¹⁰ Therefore, under exogenous stratification, the LS estimator is consistent but not asymptotically efficient.

Even under endogenous stratification, it is not difficult to obtain an estimator of θ^* that is consistent but asymptotically inefficient. To see this, assume that only Y^* is stratified. Then, by (2.6) and (3.1),

$$\begin{aligned} \mathbb{E}_{P_{Y^*|X^*}}[Y^* - \tilde{X}^{*\prime}\theta^*|X^*] = 0 \quad P_{X^*}\text{-a.s.} &\iff \mathbb{E}_{P_{Y|X}}\left[\frac{Y - \tilde{X}'\theta^*}{b(Y)}|X\right] = 0 \quad P_X\text{-a.s.} \\ &\implies \mathbb{E}_P \tilde{X} \left[\frac{Y - \tilde{X}'\theta^*}{b(Y)}\right] = 0. \end{aligned}$$

Hence, it is straightforward to show that the GMM estimator

$$\hat{\theta}_{\text{GMM, endog}} := \left(\sum_{j=1}^n \frac{\tilde{X}_j \tilde{X}_j'}{b(Y_j)} \right)^{-1} \left(\sum_{j=1}^n \frac{\tilde{X}_j Y_j}{b(Y_j)} \right)$$

is consistent for θ^* .¹¹ However, $\hat{\theta}_{\text{GMM, endog}}$ is not asymptotically efficient because its asymptotic variance is $V_{\text{GMM, endog}} := (\mathbb{E}_P \tilde{X} \tilde{X}' / b(Y))^{-1} (\mathbb{E}_{P_X} \tilde{X} \tilde{X}' V_{1, \text{endog}}(X)) (\mathbb{E}_P \tilde{X} \tilde{X}' / b(Y))^{-1}$ but

$$\text{l.b.}(\theta^* \text{ in Example 2.1})|_{\text{endog. strat.}} \leq_L V_{\text{GMM, endog}}.$$

Analogous to $\hat{\theta}_{\text{GMM, endog}}$, the GMM estimator under exogenous stratification is

$$\hat{\theta}_{\text{GMM, exog}} := \left(\sum_{j=1}^n \frac{\tilde{X}_j \tilde{X}_j'}{b(X_j)} \right)^{-1} \left(\sum_{j=1}^n \frac{\tilde{X}_j Y_j}{b(X_j)} \right),$$

which is also not asymptotically efficient because its asymptotic variance is $V_{\text{GMM, exog}} := (\mathbb{E}_{P_X} \tilde{X} \tilde{X}' / b(X))^{-1} (\mathbb{E}_{P_X} \tilde{X} \tilde{X}' V_{1, \text{exog}}(X) / b^2(X)) (\mathbb{E}_{P_X} \tilde{X} \tilde{X}' / b(X))^{-1}$ but

$$\text{l.b.}(\theta^* \text{ in Example 2.1})|_{\text{exog. strat.}} \leq_L V_{\text{GMM, exog}}.$$

Constructing efficient estimators requires more effort. For instance, suppose that stratification is purely exogenous. Then, following Robinson (1987), it can be shown that the asymptotic variance of $\hat{\theta}_{\text{Robinson}} := (\sum_{j=1}^n \tilde{X}_j \tilde{X}_j' / \hat{\sigma}^2(X_j))^{-1} (\sum_{j=1}^n \tilde{X}_j Y_j / \hat{\sigma}^2(X_j))$ equals (3.6), where $\hat{\sigma}^2$ denotes a consistent estimator of $V_{1, \text{exog}}$. Hence, $\hat{\theta}_{\text{Robinson}}$ is an asymptotically efficient estimator of θ^* under exogenous stratification. A general approach, which can be used to construct efficient estimators irrespective of whether stratification is endogenous, exogenous, or both, is discussed in Section 3.3. \square

¹⁰Namely, $M_1 \leq_L M_2$ for symmetric matrices M_1, M_2 means that $M_1 - M_2$ is negative semidefinite.

¹¹The estimator $\hat{\theta}_{\text{GMM, endog}}$ is an example of an inverse probability weighted (IPW) estimator, which uses the weights $1/b(Y_1), \dots, 1/b(Y_n)$ to correct the selection bias due to stratification by downward weighting the strata that are oversampled and upward weighting the strata that are undersampled.

Since the aggregate shares add up to one, it suffices to determine the efficiency bound for estimating $Q_{-L}^* := (Q_1^*, \dots, Q_{L-1}^*)_{(L-1) \times 1} \in (0, 1)^{L-1}$. The aggregate shares are identified in the realized population by the moment condition

$$\mathbb{E}_P\left[\frac{s(Z) - Q_{-L}^*}{b(Z)}\right] = 0, \quad (3.7)$$

where $s(Z) := (\mathbb{1}_{C_1}(Z), \dots, \mathbb{1}_{C_{L-1}}(Z))_{(L-1) \times 1}$. The moment conditions in (3.7) modify accordingly if stratification is endogenous or exogenous; namely,

$$\begin{aligned} \text{endog. strat.} &\implies \mathbb{E}_{P_Y}\left[\frac{s(Y) - Q_{-L}^*}{b(Y)}\right] = 0, & s(Y) &:= (\mathbb{1}_{A_1}(Y), \dots, \mathbb{1}_{A_{J-1}}(Y))_{(J-1) \times 1}, \\ \text{exog. strat.} &\implies \mathbb{E}_{P_X}\left[\frac{s(X) - Q_{-L}^*}{b(X)}\right] = 0, & s(X) &:= (\mathbb{1}_{B_1}(X), \dots, \mathbb{1}_{B_{M-1}}(X))_{(M-1) \times 1}. \end{aligned} \quad (3.8)$$

Let $\rho_2(Z, Q_{-L}^*) := (s(Z) - Q_{-L}^*)/b(Z)$, and $\Sigma_{12}(X) := \mathbb{E}_{P_{Y|X}}[\rho_1(Z, \theta^*)\rho_2'(Z, Q_{-L}^*)|X]$ be the conditional (on X) covariance between $\rho_1(Z, \theta^*)$ and $\rho_2(Z, Q_{-L}^*)$. Then, under (2.1), the efficiency bound for estimating Q_{-L}^* is given by

$$\begin{aligned} \text{l.b.}(Q_{-L}^*) &:= b^{*2}[\text{var}_P(\rho_2(Z, Q_{-L}^*)) - (\mathbb{E}_{P_X}\Sigma'_{12}(X)V_1^{-1}(X)\Sigma_{12}(X)) \\ &\quad + (\mathbb{E}_{P_X}\Sigma'_{12}(X)V_1^{-1}(X)D(X))(\text{l.b.}(\theta^*))(\mathbb{E}_{P_X}D'(X)V_1^{-1}(X)\Sigma_{12}(X))], \end{aligned} \quad (3.9)$$

where $\text{l.b.}(\theta^*)$ is the efficiency bound for estimating θ^* given in (3.2).

In the absence of (2.1), the efficiency bound for Q_{-L}^* is given by $b^{*2} \text{var}_P(\rho_2(Z, Q_{-L}^*))$, which follows from standard GMM theory applied to (3.7). Hence, estimating the aggregate shares in the presence of (2.1) leads to efficiency gains under endogenous stratification. There are no efficiency gains for estimating Q_{-L}^* under exogenous stratification because

$$\begin{aligned} \text{exog. strat.} &\implies \Sigma_{12}(X) = \mathbb{E}_{P_{Y|X}}\left[\frac{g(Z, \theta^*)}{b(X)} \frac{(s(X) - Q_{-L}^*)'}{b(X)}|X\right] \\ &= \mathbb{E}_{P_{Y^*|X^*}}\left[\frac{g(Z^*, \theta^*)}{b(X^*)} \frac{(s(X^*) - Q_{-L}^*)'}{b(X^*)}|X^*\right] \quad ((2.7)) \\ &= \mathbb{E}_{P_{Y^*|X^*}}[g(Z^*, \theta^*)|X^*] \frac{(s(X^*) - Q_{-L}^*)'}{b^2(X^*)} \\ &= 0 \quad P_{X^*}\text{-a.s.} \quad ((2.1)) \\ &= 0 \quad P_X\text{-a.s.} \quad (P_{X^*}^* \ll P_X \ll P_{X^*}^*) \end{aligned}$$

3.3. Efficient estimation. The estimation and testing techniques demonstrated here extend Kitamura, Tripathi, and Ahn (2004) and Tripathi and Kitamura (2003). These papers, which are based on a generalization of the empirical likelihood approach of Owen (1988), develop an asymptotically efficient methodology for estimating and testing models with conditional moment restrictions when the data are collected by random sampling.

In the papers of Kitamura, Tripathi, and Ahn, and Tripathi and Kitamura, kernel smoothing is used to efficiently incorporate the information implied by conditional moment restrictions into an empirical likelihood, which is henceforth referred to as the “smoothed empirical likelihood (SEL).” As shown in these papers, maximizing the SEL leads to one-step estimators which avoid any preliminary estimation of optimal instruments. It also yields internally studentized likelihood ratio-type statistics for testing H_0 and parametric restrictions on θ^* that do not require preliminary estimation of any variance terms. Moreover, the resulting estimation and testing procedures are invariant to normalizations of H_0 . Simulation results presented in the aforementioned papers suggest that the SEL based approach can work very well in finite samples.

The advantages of the SEL approach described above extend to the case when the observations are collected by VP sampling. Furthermore, it leads to a unified approach of estimating and testing models using stratified samples, which should appeal to applied economists and practitioners in the field. Therefore, we now demonstrate how to use the SEL approach to construct asymptotically efficient estimators, i.e., estimators with asymptotic variance equal to the efficiency bounds in Section 3.2.

If the focus is on efficient estimation of θ^* alone, then the equivalence in (3.1) reveals that replacing the moment function in Kitamura, Tripathi, and Ahn (Equation 2.1) with $\rho_1(Z, \theta^*)$ will deliver an asymptotically efficient estimator of θ^* .

But what about Q_{-L}^* ? Although the aggregate shares $Q_{-L}^* \stackrel{(3.7)}{=} \mathbb{E}_P[s(Z)]/\mathbb{E}_P[1/b(Z)]$ can be simply estimated by their sample analogs, this estimator will not be efficient because it does not take (2.1) into account; cf. the discussion after (3.9). To construct an estimator of Q_{-L}^* that accounts for (2.1), we have to jointly estimate θ^* and Q_{-L}^* , which we do using the SEL approach.

For the remainder of the paper, assume that we have independent observations Z_1, \dots, Z_n collected by VP sampling. Hence, these are i.i.d. draws from the realized density dP in (2.3). Let $X := (X^c, X^d)$, where X^c denotes the vector of continuous components and X^d the vector of discrete components. Our estimation approach relies on a smoothed version of empirical likelihood. This smoothing, or localization, is carried out using positive kernel weights

$$w_{ij} := \frac{\mathcal{K}_{b_n}(X_i^c - X_j^c) \mathbb{1}(X_i^d = X_j^d)}{\sum_{k=1}^n \mathcal{K}_{b_n}(X_i^c - X_k^c) \mathbb{1}(X_i^d = X_k^d)}, \quad i, j = 1, \dots, n,$$

where \mathcal{K} is a second order kernel, $\mathcal{K}_{b_n}(\cdot) := \mathcal{K}(\cdot/b_n)$, and the bandwidth b_n is a null sequence of positive numbers such that $nb_n^{\dim(X^c)} \rightarrow \infty$ as $n \rightarrow \infty$.

For $i, j = 1, \dots, n$, let p_{ij} denote the probability mass placed at (X_i, Z_j) by a discrete distribution with support $(X_1, \dots, X_n) \times (Z_1, \dots, Z_n)$. The collection of probabilities $(p_{ij})_{i,j=1}^n$ can be thought of as a set of nuisance parameters that includes the empirical distribution of

the data. Using the kernel weights (w_{ij}) and the distribution (p_{ij}) construct the smoothed loglikelihood $\sum_{i=1}^n \sum_{j=1}^n w_{ij} \log p_{ij}$. Then, given (θ, Q_{-L}) , concentrate out (p_{ij}) by solving the following optimization problem:

$$\begin{aligned} & \max_{(p_{ij})} \sum_{i=1}^n \sum_{j=1}^n w_{ij} \log p_{ij} \\ & \text{s.t. } p_{ij} \geq 0 \text{ for } i, j = 1, \dots, n, \quad \sum_{i=1}^n \sum_{j=1}^n p_{ij} = 1, \\ & \sum_{j=1}^n \rho_1(Z_j, \theta) p_{1j} = 0, \dots, \sum_{j=1}^n \rho_1(Z_j, \theta) p_{nj} = 0, \quad \sum_{i=1}^n \sum_{j=1}^n \rho_2(Z_j, Q_{-L}) p_{ij} = 0. \end{aligned} \quad (3.10)$$

If the convex hulls of $\{\rho_1(Z_1, \theta), \dots, \rho_1(Z_n, \theta)\}$ and $\{\rho_2(Z_1, Q_{-L}), \dots, \rho_2(Z_n, Q_{-L})\}$ contain the origin, then (3.10) can be solved by using Lagrange multipliers. In this case, it can be verified that the solution to (3.10) is given by

$$\hat{p}_{ij}(\theta, Q_{-L}) := \frac{1}{n} \left(\frac{w_{ij}}{1 + \lambda'_i \rho_1(Z_j, \theta) + \mu' \rho_2(Z_j, Q_{-L})} \right), \quad i, j = 1, \dots, n,$$

where the multipliers $\lambda_i := \lambda_i(\theta, Q_{-L})$ and $\mu := \mu(\theta, Q_{-L})$ solve

$$\begin{aligned} & \sum_{j=1}^n \frac{w_{ij} \rho_1(Z_j, \theta)}{1 + \lambda'_i \rho_1(Z_j, \theta) + \mu' \rho_2(Z_j, Q_{-L})} = 0, \quad i = 1, \dots, n, \\ & \sum_{i=1}^n \sum_{j=1}^n \frac{w_{ij} \rho_2(Z_j, Q_{-L})}{1 + \lambda'_i \rho_1(Z_j, \theta) + \mu' \rho_2(Z_j, Q_{-L})} = 0. \end{aligned} \quad (3.11)$$

The smoothed empirical loglikelihood of (θ, Q_{-L}) is given by

$$\begin{aligned} \text{SEL}(\theta, Q_{-L}) & := \sum_{i=1}^n \sum_{j=1}^n w_{ij} \log \hat{p}_{ij}(\theta, Q_{-L}) \\ & = \sum_{i=1}^n \sum_{j=1}^n w_{ij} \log \left(\frac{w_{ij}/n}{1 + \lambda'_i \rho_1(Z_j, \theta) + \mu' \rho_2(Z_j, Q_{-L})} \right), \end{aligned} \quad (3.12)$$

where the multipliers solve (3.11).

The estimators of θ^* and Q_{-L}^* can, in principle, be defined to be the maximizers of $\text{SEL}(\theta, Q_{-L})$. However, this leads to a constrained optimization problem because the Lagrange multipliers in $\text{SEL}(\theta, Q_{-L})$ have to satisfy (3.11). To ease computation, it is better to convert the constrained optimization problem into an unconstrained optimization problem as follows. Begin by observing that, by (3.12),

$$\text{SEL}(\theta, Q_{-L}) = \sum_{i=1}^n \sum_{j=1}^n w_{ij} \log(w_{ij}/n) - \sum_{i=1}^n \sum_{j=1}^n w_{ij} \log(1 + \lambda'_i \rho_1(Z_j, \theta) + \mu' \rho_2(Z_j, Q_{-L})).$$

Furthermore, comparing (3.11) with the first order conditions for (3.13), it is straightforward to see that

$$(\lambda_1, \dots, \lambda_n, \mu) = \operatorname{argmax}_{\tilde{\lambda}_1, \dots, \tilde{\lambda}_n, \tilde{\mu}} \sum_{i=1}^n \sum_{j=1}^n w_{ij} \log(1 + \tilde{\lambda}'_i \rho_1(Z_j, \theta) + \tilde{\mu}' \rho_2(Z_j, Q_{-L})). \quad (3.13)$$

Therefore, the estimators of θ^* and Q_{-L}^* are defined to be

$$(\hat{\theta}, \hat{Q}_{-L}) := \operatorname{argmax}_{\theta, Q_{-L}} \operatorname{SEL}_{\mathbb{T}}(\theta, Q_{-L}), \quad (3.14)$$

where the “trimmed” SEL objective function

$$\begin{aligned} \operatorname{SEL}_{\mathbb{T}}(\theta, Q_{-L}) &:= - \max_{\tilde{\lambda}_1, \dots, \tilde{\lambda}_n, \tilde{\mu}} \sum_{i=1}^n \mathbb{T}_{i,n} \sum_{j=1}^n w_{ij} \log(1 + \tilde{\lambda}'_i \rho_1(Z_j, \theta) + \tilde{\mu}' \rho_2(Z_j, Q_{-L})) \\ &= - \max_{\tilde{\mu}} \sum_{i=1}^n \mathbb{T}_{i,n} \max_{\tilde{\lambda}_i} \sum_{j=1}^n w_{ij} \log(1 + \tilde{\lambda}'_i \rho_1(Z_j, \theta) + \tilde{\mu}' \rho_2(Z_j, Q_{-L})). \end{aligned} \quad (3.15)$$

The trimming indicator $\mathbb{T}_{i,n} := \mathbb{1}(\hat{h}(X_i) \geq b_n^\tau)$, where $\hat{h}(X_i) := \sum_{j=1}^n \mathcal{K}_{b_n}(X_i^c - X_j^c) \mathbb{1}(X_i^d = X_j^d) / (nb_n^{\dim(X^c)})$ and $\tau \in (0, 1)$ is a trimming parameter, is incorporated in (3.15) to deal with the “denominator problem,” namely, the instability of the local empirical loglikelihood $\sum_{j=1}^n w_{ij} \log(1 + \tilde{\lambda}'_i \rho_1(Z_j, \theta) + \tilde{\mu}' \rho_2(Z_j, Q_{-L}))$ caused by the density of the conditioning variables becoming too small in the tails. Since $\mathbb{T}_{i,n} \xrightarrow{P} 1$ as $n \rightarrow \infty$, this trimming scheme ensures that asymptotically no data is lost.

Following Kitamura, Tripathi, and Ahn, it can be shown that, under some regularity conditions, $\hat{\theta}$ and \hat{Q}_{-L} are consistent, asymptotically normal, and asymptotically efficient, i.e., their asymptotic variances match the efficiency bounds.

3.4. Testing. The empirical likelihood approach provides a convenient unified environment for testing. For instance, suppose we want to test the parametric restriction $\tilde{H}_0 : R(\theta^*) = 0$ against the alternative that \tilde{H}_0 is false, where R is a vector of twice continuously differentiable functions such that $\operatorname{rank} \partial_\theta R(\theta^*) = \dim(R)$. Let

$$(\hat{\theta}_R, \hat{Q}_{-L,R}) := \operatorname{argmax}_{\theta, Q_{-L}} \operatorname{SEL}_{\mathbb{T}}(\theta, Q_{-L}) \quad \text{s.t.} \quad R(\theta) = 0.$$

A version of the likelihood ratio statistic for testing \tilde{H}_0 is given by

$$\operatorname{LR} := 2[\operatorname{SEL}_{\mathbb{T}}(\hat{\theta}, \hat{Q}_{-L}) - \operatorname{SEL}_{\mathbb{T}}(\hat{\theta}_R, \hat{Q}_{-L,R})].$$

It can be shown that, under some regularity conditions, $\operatorname{LR} \xrightarrow[n \rightarrow \infty]{d} \chi_{\dim(R)}^2$ whenever \tilde{H}_0 is true. This result can be used to obtain the critical values for LR. Although a Wald statistic can also be constructed, it is less attractive than LR because the latter is internally studentized. As in parametric situations, LR can be inverted to obtain asymptotically valid confidence intervals. A nice property of confidence intervals based on LR is that they are invariant to

nonsingular transformations of the moment conditions. They also automatically satisfy natural range restrictions.

Since inference based on $\hat{\theta}$ is sensible only if (2.1) is true, it is important to devise a test for H_0 against the alternative that it is false. As we are dealing with conditional moment restrictions, any specification test which first converts (2.1) into a finite set of unconditional moment restrictions will not be consistent for testing H_0 . However, using the equivalence in (3.1), a consistent test of H_0 is easily obtained by replacing moment function in Tripathi and Kitamura (2003, Equation 1.1) with $\rho_1(Z, \theta^*)$. Note that since (3.7) just identifies the aggregate shares, testing the specification of (2.1) and (3.7) jointly is equivalent to testing (2.1).

4. SIMULATION STUDY

We now examine the finite sample behavior of the LS, GMM, and SEL estimators to illustrate the effects of estimating a simple linear regression model specified for the target population, when data is collected by VP sampling and stratification is either endogenous or exogenous. The SEL estimator of the model parameters and aggregate shares defined in (3.14) is implemented using the algorithm in Owen (2013). Code for the simulations is written in R.

4.1. Design. We consider the design in Kitamura, Tripathi, and Ahn (Section 5), which has been used earlier by Cragg (1983) and Newey (1993). The model to be estimated is

$$Y^* = \beta_0^* + \beta_1^* X^* + \sigma^*(X^*) \varepsilon^*, \quad (4.1)$$

where $\mathbb{E}_{P_{Y^*|X^*}}[\varepsilon^*|X^*] = 0$ P_{X^*} -a.s., $\theta^* := (\beta_0^*, \beta_1^*) = (1, 1)$, and $(\varepsilon^*, \log X^*) \stackrel{d}{=} \text{NIID}(0, 1)$. We consider two specifications for the scedastic function in the target population: a (conditional) heteroscedastic design, relevant for applications, with $\sigma^*(X^*) := \sqrt{0.1 + 0.2X^* + 0.3X^{*2}}$; and a (conditional) homoscedastic design, essentially of theoretical interest, with $\sigma^*(X^*) := 1$.

The target population is partitioned into two strata. Under endogenous stratification, $\mathbb{A}_1 = (-\infty, 1.4)$ and $\mathbb{A}_2 = [1.4, \infty)$. Under exogenous stratification, $\mathbb{B}_1 = \mathbb{A}_1$ and $\mathbb{B}_2 = \mathbb{A}_2$. The aggregate shares for the four configurations are given in Table 1. The VP sampling probabilities are $(p_1, p_2) = (0.9, 0.3)$; i.e., the first stratum is heavily oversampled, irrespective of whether the stratification is endogenous or exogenous. Since it is typically strata with small aggregate shares that are oversampled, this sampling design focuses on endogenous stratification, which is the object of attention in most applications.

Tables 2 and 3 reports the summary statistics averaged across 1000 Monte Carlo replications for the LS estimator $\hat{\theta}_{\text{LS}}$, the GMM estimators $\hat{\theta}_{\text{GMM, endog}}$ and $\hat{\theta}_{\text{GMM, exog}}$, and the SEL estimator $\hat{\theta}$.¹² Three sample sizes are considered, namely, $n = 50, 150, 500$. Tables 4 and 5,

¹²The SEL estimator is implemented with $\mathbb{T}_{i,n} := 1$. To the best of our knowledge, how to choose an optimal data driven bandwidth for the SEL estimator remains an open problem. Consequently, we naively chose the bandwidth by repeating the simulation experiment on a grid of bandwidths and picking the one that

TABLE 1. Aggregate shares for the simulation study.

Stratification	Design	Q_1^*
endogenous	homoscedastic	0.27
	heteroscedastic	0.28
exogenous	homoscedastic	0.63
	heteroscedastic	0.63

which summarize the simulation results for estimating Q_1^* , compare the GMM estimators based on the moment conditions in (3.8) with the SEL estimator \hat{Q}_1 .

4.2. Discussion. Recall that the LS estimator is inconsistent under endogenous stratification, and consistent but inefficient under exogenous stratification; the GMM estimators are consistent but inefficient under endogenous and exogenous stratification; and the SEL estimator is consistent and asymptotically efficient irrespective of whether the stratification is endogenous or exogenous. Tables 2–5 largely confirm these results, at least as far as estimating the model parameters is concerned.

The inconsistency of the LS estimator of β_1^* under endogenous stratification is apparent from Tables 2 and 3 because the bias of the LS estimator, as a fraction of β_1^* , remains greater than 9% in magnitude under heteroscedasticity, and greater than 6% under homoscedasticity, as the sample size increases from 50 to 500.¹³ In contrast, in both designs, the LS and GMM estimators under exogenous stratification are practically unbiased even when $n = 50$. Under exogenous stratification, the LS estimator has smaller sampling variance than the GMM estimator for each sample size. However, this finding can be mathematically justified only for homoscedastic designs. Indeed, as shown in the appendix, cf. Example A.1, counterexamples can be constructed to show that in heteroscedastic designs the LS estimator can have higher

minimized the average (across the simulation replications) RMSE of the SEL estimator of β_1^* . The naively chosen bandwidth, labeled c_n , is reported in Tables 2–5. For the sake of comparison, we also report the SEL estimator when the bandwidth is chosen using Silverman’s rule of thumb, namely, $b_n = 1.06 \widehat{\text{sd}}(X) n^{-1/5}$. Since $\widehat{\text{sd}}(X)$ depends on the data, the b_n reported in the tables is the value averaged across the simulations.

¹³This is even more so for the LS estimator of the intercept because, under endogenous stratification, the bias of the LS estimator of β_0^* is $\approx 18\%$ (resp. $\approx 41\%$) in magnitude for the heteroscedastic (resp. homoscedastic) design even when $n = 500$. For the remainder of this section, however, we only discuss the simulation results for the slope coefficient because it can be interpreted as an average partial effect. Results for the intercept, which is a pure level effect, are qualitatively very similar.

sampling variance than the GMM estimator when stratification is exogenous.¹⁴ Under endogenous stratification, the GMM estimator of the slope coefficient exhibits some bias ($\approx 2\text{--}4\%$ in both designs) when $n = 50$, but the bias is very close to zero when $n = 500$. This is true whether the design is homoscedastic or heteroscedastic, although the magnitude of the bias is higher under heteroscedasticity.

Tables 2–5 reveal that the SEL estimator using the naively chosen bandwidth (c_n) behaves very similarly to the SEL estimator using the Silverman’s rule of thumb bandwidth (b_n). Hence, subsequent discussion regarding the SEL estimator is based on its implementation using the naively chosen bandwidth.

The SEL estimator of β_1^* is consistent whether stratification is exogenous or endogenous. In the heteroscedastic design, the SEL estimator exhibits some bias ($\approx 1\%$) under endogenous stratification when $n = 500$, although its bias under exogenous stratification is close to zero. Moreover, in the heteroscedastic design, the SEL estimator beats the GMM estimator in terms of the RMSE under each stratification scheme and for each sample size. Not surprisingly, the contrast between the two is most pronounced when $n = 500$; e.g., irrespective of the stratification scheme, the RMSE of the GMM estimator is at least 65% larger than the RMSE of the SEL estimator.

In the homoscedastic design, even though it exhibits some bias under endogenous and exogenous stratification when $n = 50$, the bias of the SEL estimator is close to zero for $n = 500$. However, its RMSE is larger than that of the GMM estimator even when $n = 500$. This finding, which corroborates the simulation results in Kitamura, Tripathi, and Ahn (p. 1682), is likely due to the fact that the SEL estimator internally estimates the scedastic function nonparametrically to achieve semiparametric efficiency and is thus unable to take advantage of conditional homoscedasticity in small samples.

Tables 4 and 5 reveal that the GMM estimator of Q_1^* is consistent whether stratification is endogenous or exogenous. It exhibits some upward bias ($\approx 1\text{--}2\%$) in both designs and for both types of stratification when $n = 50$, but the bias is very close to zero when $n = 500$.¹⁵ In both designs, the RMSE of the SEL estimator of Q_1^* is always slightly larger than the RMSE of the of the GMM estimator under endogenous stratification, implying that in small samples

¹⁴It is shown in the appendix, cf. (A.1), that $\text{asvar}(n^{1/2}(\hat{\theta}_{\text{GMM,exog}} - \theta^*)) - \text{asvar}(n^{1/2}(\hat{\theta}_{\text{LS}} - \theta^*)) = A + B$ holds under exogenous stratification, where the matrix A is positive semidefinite and the matrix B is negative semidefinite. Therefore, in general, it is not clear which estimator has smaller asymptotic variance. However, since $B = 0$ under homoscedasticity, cf. (A.3), $\text{asvar}(n^{1/2}(\hat{\theta}_{\text{LS}} - \theta^*)) \leq_L \text{asvar}(n^{1/2}(\hat{\theta}_{\text{GMM,exog}} - \theta^*))$ holds under exogenous stratification and homoscedasticity. Alternatively, under homoscedasticity, the Gauss-Markov theorem implies the same result because $\hat{\theta}_{\text{GMM,exog}}$ and $\hat{\theta}_{\text{LS}}$ are both linear unbiased when stratification is exogenous.

¹⁵In Tables 4 and 5, the results under exogenous stratification are almost identical for the heteroscedastic and homoscedastic designs because $P^*(X^* \in \mathbb{B}_1)$ is not affected by conditional heteroscedasticity in Y^* (cf. Table 1).

there appears to be no efficiency gain in estimating Q_1^* jointly with the model parameters. As can be seen from Tables 4 and 5, the increase in the RMSE of \hat{Q}_1 is due to its bias, because $\text{RMSE} \approx \text{SE}$ whenever the bias is small. This becomes clear on comparing the bias of \hat{Q}_1 under endogenous and exogenous stratification: the latter is always larger. The higher bias of \hat{Q}_1 under exogenous stratification is likely a design effect.

5. CONCLUSION

When estimating or testing economic relationships, economists often discover that the data they plan to use is not drawn randomly from the target population for which they wish to draw an inference. Instead, the observations are found to be sampled from a related but different distribution. If this feature is not taken into account when doing statistical analysis, subsequent inference can be severely biased. In this paper, we show how to use a smoothed empirical likelihood approach to conduct efficient semiparametric inference in models characterized as conditional moment equalities when data is collected by variable probability sampling. Results from a simulation experiment suggest that the smoothed empirical likelihood based estimator can estimate the model parameters very well in small to moderately sized stratified samples.

ACKNOWLEDGEMENTS

The simulation experiments reported in this paper were carried out using the HPC facilities of the University of Luxembourg (Varrette et al., 2014, <http://hpc.uni.lu>).

APPENDIX A. LS VERSUS GMM UNDER EXOGENOUS STRATIFICATION

Recalling the expressions for $V_{\text{GMM,exog}}$ and $V_{\text{LS,exog}}$ in Example 3.1, we can write

$$\text{asvar}(n^{1/2}(\hat{\theta}_{\text{GMM,exog}} - \theta^*)) - \text{asvar}(n^{1/2}(\hat{\theta}_{\text{LS}} - \theta^*)) = \left(\mathbb{E}_{P_X} \frac{\tilde{X}\tilde{X}'}{b(X)} \right)^{-1} \Omega \left(\mathbb{E}_{P_X} \frac{\tilde{X}\tilde{X}'}{b(X)} \right)^{-1},$$

where

$$\begin{aligned} \Omega := & \left(\mathbb{E}_{P_X} \tilde{X}\tilde{X}' \frac{V_{1,\text{exog}}(X)}{b^2(X)} \right) \\ & - \left(\mathbb{E}_{P_X} \frac{\tilde{X}\tilde{X}'}{b(X)} \right) \left(\mathbb{E}_{P_X} \tilde{X}\tilde{X}' \right)^{-1} \left(\mathbb{E}_{P_X} \tilde{X}\tilde{X}' V_{1,\text{exog}}(X) \right) \left(\mathbb{E}_{P_X} \tilde{X}\tilde{X}' \right)^{-1} \left(\mathbb{E}_{P_X} \frac{\tilde{X}\tilde{X}'}{b(X)} \right). \end{aligned}$$

Next, letting $a_1 := \tilde{X}V_{1,\text{exog}}(X)/b(X)$ and $a_2 := (\mathbb{E}_{P_X} \tilde{X}\tilde{X}')^{-1} \tilde{X}/V_{1,\text{exog}}(X)$, we have that

$$\begin{aligned} \Omega &= \mathbb{E}_{P_X} a_1 a_1' - (\mathbb{E}_{P_X} a_1 a_2') (\mathbb{E}_{P_X} \tilde{X}\tilde{X}' V_{1,\text{exog}}(X)) (\mathbb{E}_{P_X} a_2 a_1') \\ &= \mathbb{E}_{P_X} a_1 a_1' - (\mathbb{E}_{P_X} a_1 a_2') (\mathbb{E}_{P_X} a_2 a_2')^{-1} (\mathbb{E}_{P_X} a_2 a_1') \\ &\quad + (\mathbb{E}_{P_X} a_1 a_2') [(\mathbb{E}_{P_X} a_2 a_2')^{-1} - (\mathbb{E}_{P_X} \tilde{X}\tilde{X}' V_{1,\text{exog}}(X))] (\mathbb{E}_{P_X} a_2 a_1'). \end{aligned}$$

Consequently, under exogenous stratification we can write

$$\text{asvar}(n^{1/2}(\hat{\theta}_{\text{GMM,exog}} - \theta^*)) - \text{asvar}(n^{1/2}(\hat{\theta}_{\text{LS}} - \theta^*)) = A + B, \quad (\text{A.1})$$

where

$$A := \left(\mathbb{E}_{P_X} \frac{\tilde{X}\tilde{X}'}{b(X)} \right)^{-1} [\mathbb{E}_{P_X} a_1 a_1' - (\mathbb{E}_{P_X} a_1 a_2') (\mathbb{E}_{P_X} a_2 a_2')^{-1} (\mathbb{E}_{P_X} a_2 a_1')] \left(\mathbb{E}_{P_X} \frac{\tilde{X}\tilde{X}'}{b(X)} \right)^{-1}$$

$$B := \left(\mathbb{E}_{P_X} \frac{\tilde{X}\tilde{X}'}{b(X)} \right)^{-1} (\mathbb{E}_{P_X} a_1 a_2') [(\mathbb{E}_{P_X} a_2 a_2')^{-1} - (\mathbb{E}_{P_X} \tilde{X}\tilde{X}' V_{1,\text{exog}}(X))] \left(\mathbb{E}_{P_X} \frac{\tilde{X}\tilde{X}'}{b(X)} \right)^{-1} (\mathbb{E}_{P_X} a_2 a_1').$$

It remains to show that A is positive semidefinite and B is negative semidefinite. To do so, recall the matrix version of the Cauchy-Schwarz inequality (Tripathi, 1999), namely,

$$(\mathbb{E}GH')(\mathbb{E}HH')^{-1}(\mathbb{E}HG') \leq_L \mathbb{E}GG', \quad (\text{A.2})$$

where G and H are random column vectors. Then, letting $G := a_1$ and $H := a_2$, it is immediate from (A.2) that A is positive semidefinite. Next,

$$(\mathbb{E}_{P_X} a_2 a_2')^{-1} = (\mathbb{E}_{P_X} \tilde{X}\tilde{X}') \left(\mathbb{E}_{P_X} \frac{\tilde{X}\tilde{X}'}{V_{1,\text{exog}}(X)} \right)^{-1} (\mathbb{E}_{P_X} \tilde{X}\tilde{X}')$$

$$\leq_L \mathbb{E}_{P_X} \tilde{X}\tilde{X}' V_{1,\text{exog}}(X)$$

follows from (A.2) on letting $G := \tilde{X} \sqrt{V_{1,\text{exog}}(X)}$ and $H := \tilde{X} / \sqrt{V_{1,\text{exog}}(X)}$. Hence, B is negative semidefinite. Consequently, as A is positive semidefinite and B is negative semidefinite, it is not clear from (A.1) which estimator has smaller asymptotic variance.

However, if conditional homoscedasticity holds in the target population, then

$$\text{var}_{P^*}(Y^*|X^*) = \text{const.} \quad P_{X^*}^*\text{-a.s.}$$

Moreover, under exogenous stratification,

$$\text{var}_{P^*}(Y^*|X^* = x) \stackrel{(2.7)}{=} \text{var}_P(Y|X = x), \quad x \in \text{supp}(X^*).$$

Hence, since $P_{X^*}^* \ll P_X \ll P_{X^*}^*$, conditional homoscedasticity and exogenous stratification together imply that

$$V_{1,\text{exog}}(X) = \text{var}_P(Y|X) = \text{const.} \quad P_X\text{-a.s.}$$

Therefore, under conditional homoscedasticity and exogenous stratification,

$$(\mathbb{E}_{P_X} a_2 a_2')^{-1} - (\mathbb{E}_{P_X} \tilde{X}\tilde{X}' V_{1,\text{exog}}(X))$$

$$= (\mathbb{E}_{P_X} \tilde{X}\tilde{X}') \left(\mathbb{E}_{P_X} \frac{\tilde{X}\tilde{X}'}{V_{1,\text{exog}}(X)} \right)^{-1} (\mathbb{E}_{P_X} \tilde{X}\tilde{X}') - (\mathbb{E}_{P_X} \tilde{X}\tilde{X}' V_{1,\text{exog}}(X))$$

$$= 0.$$

It follows from the definition of B that

$$\text{conditional homoscedasticity and exogenous stratification} \implies B = 0. \quad (\text{A.3})$$

Hence, $\text{asvar}(n^{1/2}(\hat{\theta}_{\text{LS}} - \theta^*)) \leq_L \text{asvar}(n^{1/2}(\hat{\theta}_{\text{GMM,exog}} - \theta^*))$ holds under exogenous stratification and conditional homoscedasticity.

However, as demonstrated in the next example, this result may not hold under conditional heteroscedasticity.

Example A.1. Consider (4.1) with $\beta_0 := 0$, i.e., a simple linear regression through the origin.

As before, $\mathbb{E}_{P_{Y^*|X^*}}[\varepsilon^*|X^*] = 0$ P_{X^*} -a.s.. Assume that only $X^* := \begin{cases} c & \text{w.p. } 1-r \\ d & \text{w.p. } r \end{cases}$ is stratified

with $L = 2$, where $\mathbb{B}_1 = (-\infty, 0)$ and $\mathbb{B}_2 = [0, +\infty)$.

Under exogenous stratification,

$$\begin{aligned} \text{asvar}(n^{1/2}(\hat{\beta}_{1,\text{LS}} - \beta_1^*)) &= \frac{\mathbb{E}_{P_X} X^2 V_{1,\text{exog}}(X)}{(\mathbb{E}_{P_X} X^2)^2} = \frac{\mathbb{E}_{P_X} X^2 \sigma^{*2}(X)}{(\mathbb{E}_{P_X} X^2)^2} \\ \text{asvar}(n^{1/2}(\hat{\beta}_{1,\text{GMM}} - \beta_1^*)) &= \frac{\mathbb{E}_{P_X} [X^2 V_{1,\text{exog}}(X)/b^2(X)]}{(\mathbb{E}_{P_X} X^2/b(X))^2} = \frac{\mathbb{E}_{P_X} [X^2 \sigma^{*2}(X)/b^2(X)]}{(\mathbb{E}_{P_X} X^2/b(X))^2}. \end{aligned}$$

Let $r = 1/3$, $c = -1$, $d = 2$, $\sigma^{*2}(c) = 1$, $\sigma^{*2}(d) = 4$, $p_1 = 0.2$, and $p_2 = 0.8$. Note that $b(c) = p_1 \mathbb{1}_{\mathbb{B}_1}(c) + p_2 \mathbb{1}_{\mathbb{B}_2}(c) = p_1$ because $c < 0$, and $b(d) = p_1 \mathbb{1}_{\mathbb{B}_1}(d) + p_2 \mathbb{1}_{\mathbb{B}_2}(d) = p_2$ because $d > 0$. Then, it can be verified that

$$\mathbb{E}_{P_X} X^2 \sigma^{*2}(X) = 6, \quad \mathbb{E}_{P_X} X^2 = 2, \quad \mathbb{E}_{P_X} [X^2 \sigma^{*2}(X)/b^2(X)] = 25, \quad \mathbb{E}_{P_X} [X^2/b(X)] = 5.$$

Consequently,

$$\text{asvar}(n^{1/2}(\hat{\beta}_{1,\text{LS}} - \beta_1^*)) = 1.5 \quad > \quad \text{asvar}(n^{1/2}(\hat{\beta}_{1,\text{GMM}} - \beta_1^*)) = 1.$$

This shows that the LS estimator may be asymptotically inefficient compared to the GMM estimator under conditional heteroscedasticity and exogenous stratification. \square

TABLE 2. Simulation summary: Estimated β_0^*, β_1^* under heteroscedasticity.

Stratification	n	Estimator	Intercept			Slope		
			Bias	SE	RMSE	Bias	SE	RMSE
endogenous	50	LS	-.1595	.4417	.4694	-.1076	.4868	.4983
		GMM	.0307	.5165	.5171	-.0408	.4757	.4772
		SEL ($c_n = 0.3$)	-.0033	.2909	.2910	-.0329	.3845	.3859
		SEL ($b_n \approx 0.46$)	.0015	.3916	.3916	-.0213	.4140	.4146
	150	LS	-.1714	.3489	.3885	-.1025	.3514	.3658
		GMM	.0234	.3980	.3985	-.0248	.3325	.3332
		SEL ($c_n = 0.4$)	.0202	.1880	.1891	-.0332	.2394	.2417
		SEL ($b_n \approx 0.39$)	.0224	.2313	.2324	-.0296	.2543	.2560
	500	LS	-.1805	.2894	.3410	-.0906	.2641	.2790
		GMM	.0043	.3304	.3302	-.0061	.2456	.2456
		SEL ($c_n = 0.8$)	.0107	.1316	.1321	-.0130	.1486	.1492
		SEL ($b_n \approx 0.31$)	.0096	.1242	.1246	-.0131	.1454	.1460
exogenous	50	LS	.0038	.3435	.3434	-.0032	.4275	.4273
		GMM	.0080	.4863	.4861	-.0041	.4791	.4789
		SEL ($c_n = 0.3$)	-.0161	.2518	.2523	.0250	.3754	.3762
		SEL ($b_n \approx 0.29$)	-.0098	.3547	.3549	.0117	.4225	.4227
	150	LS	.0021	.2609	.2608	-.0063	.3062	.3061
		GMM	.0042	.3838	.3836	-.0070	.3364	.3363
		SEL ($c_n = 0.4$)	.0010	.1562	.1562	-.0026	.2326	.2326
		SEL ($b_n \approx 0.24$)	.0020	.1910	.1910	-.0034	.2472	.2472
	500	LS	-.0012	.2189	.2188	.0014	.2323	.2322
		GMM	-.0023	.3354	.3352	.0012	.2540	.2539
		SEL ($c_n = 0.8$)	.0006	.1200	.1200	.0017	.1530	.1530
		SEL ($b_n \approx 0.19$)	.0003	.0988	.0988	.0017	.1425	.1425

TABLE 3. Simulation summary: Estimated β_0^*, β_1^* under homoscedasticity.

Stratification	n	Estimator	Intercept			Slope		
			Bias	SE	RMSE	Bias	SE	RMSE
endogenous	50	LS	-.4576	.2855	.5392	.0991	.2156	.2372
		GMM	-.0431	.3389	.3415	.0158	.2231	.2236
		SEL ($c_n = 0.3$)	-.0790	.4180	.4255	.0379	.3501	.3521
		SEL ($b_n \approx 0.42$)	-.0559	.3766	.3807	.0255	.2793	.2805
	150	LS	-.4273	.1480	.4522	.0708	.0906	.1149
		GMM	-.0053	.1680	.1680	-.0011	.0902	.0902
		SEL ($c_n = 0.4$)	-.0160	.2028	.2034	.0070	.1364	.1366
		SEL ($b_n \approx 0.35$)	-.0135	.1914	.1919	.0069	.1235	.1237
	500	LS	-.4142	.0845	.4227	.0626	.0453	.0772
		GMM	-.0005	.0938	.0938	.0000	.0427	.0427
		SEL ($c_n = 0.8$)	-.0047	.1026	.1027	.0031	.0577	.0578
		SEL ($b_n \approx 0.28$)	-.0061	.1053	.1055	.0037	.0590	.0591
exogenous	50	LS	-.0039	.2432	.2431	.0031	.1984	.1983
		GMM	-.0022	.2622	.2621	.0022	.2084	.2083
		SEL ($c_n = 0.3$)	-.0184	.3153	.3159	.0230	.3214	.3222
		SEL ($b_n \approx 0.29$)	-.0077	.2958	.2959	.0078	.2787	.2788
	150	LS	-.0007	.1260	.1260	-.0022	.0843	.0843
		GMM	.0011	.1314	.1314	-.0024	.0863	.0863
		SEL ($c_n = 0.4$)	.0001	.1502	.1502	-.0007	.1261	.1261
		SEL ($b_n \approx 0.24$)	.0001	.1516	.1516	-.0012	.1199	.1199
	500	LS	.0027	.0703	.0703	-.0006	.0410	.0410
		GMM	.0040	.0755	.0756	-.0008	.0416	.0416
		SEL ($c_n = 0.8$)	.0053	.0785	.0787	-.0029	.0554	.0555
		SEL ($b_n \approx 0.19$)	.0033	.0801	.0802	-.0012	.0595	.0595

TABLE 4. Simulation summary: Estimated Q_1^* under heteroscedasticity.

Stratification	Sample size	Estimator	Bias	SE	RMSE
endogenous	50	GMM	.0132	.0890	.0900
		SEL ($c_n = 0.3$)	.0178	.0939	.0956
		SEL ($b_n \approx 0.46$)	.0209	.0940	.0963
	150	GMM	.0047	.0504	.0506
		SEL ($c_n = 0.4$)	.0096	.0532	.0540
		SEL ($b_n \approx 0.39$)	.0126	.0534	.0549
	500	GMM	.0014	.0278	.0278
		SEL ($c_n = 0.8$)	.0106	.0294	.0313
		SEL ($b_n \approx 0.31$)	.0092	.0293	.0307
exogenous	50	GMM	.0133	.1070	.1078
		SEL ($c_n = 0.3$)	.0384	.1102	.1167
		SEL ($b_n \approx 0.29$)	.0550	.1032	.1169
	150	GMM	.0050	.0633	.0635
		SEL ($c_n = 0.4$)	.0414	.0655	.0775
		SEL ($b_n \approx 0.24$)	.0523	.0620	.0811
	500	GMM	.0009	.0347	.0347
		SEL ($c_n = 0.8$)	.0719	.0364	.0806
		SEL ($b_n \approx 0.19$)	.0471	.0344	.0583

TABLE 5. Simulation summary: Estimated Q_1^* under homoscedasticity.

Stratification	Sample size	Estimator	Bias	SE	RMSE
endogenous	50	GMM	.0135	.0873	.0883
		SEL ($c_n = 0.3$)	.0204	.0909	.0931
		SEL ($b_n \approx 0.42$)	.0266	.0924	.0962
	150	GMM	.0042	.0492	.0493
		SEL ($c_n = 0.4$)	.0129	.0515	.0531
		SEL ($b_n \approx 0.35$)	.0176	.0519	.0548
	500	GMM	.0013	.0262	.0263
		SEL ($c_n = 0.8$)	.0180	.0284	.0337
		SEL ($b_n \approx 0.28$)	.0150	.0284	.0321
exogenous	50	GMM	.0133	.1070	.1078
		SEL ($c_n = 0.3$)	.0380	.1088	.1153
		SEL ($b_n \approx 0.29$)	.0561	.1020	.1164
	150	GMM	.0050	.0633	.0635
		SEL ($c_n = 0.4$)	.0420	.0654	.0777
		SEL ($b_n \approx 0.24$)	.0530	.0619	.0815
	500	GMM	.0009	.0347	.0347
		SEL ($c_n = 0.8$)	.0720	.0364	.0807
		SEL ($b_n \approx 0.19$)	.0473	.0344	.0585

REFERENCES

- BHATTACHARYA, D. (2005): "Asymptotic inference from multi-stage samples," *Journal of Econometrics*, 126, 145–171.
- (2007): "Inference on inequality from household survey data," *Journal of Econometrics*, 137, 674–707.
- BICKEL, P. J., C. A. J. KLASSEN, Y. RITOV, AND J. A. WELLNER (1993): *Efficient and adaptive estimation for semiparametric models*. Johns Hopkins Press.
- BICKEL, P. J., AND Y. RITOV (1991): "Large sample theory of estimation in biased sampling regression models," *Annals of Statistics*, 19, 797–816.
- BUTLER, J. S. (2000): "Efficiency results of MLE and GMM estimation with sampling weights," *Journal of Econometrics*, 96, 25–37.
- CHAMBERLAIN, G. (1987): "Asymptotic efficiency in estimation with conditional moment restrictions," *Journal of Econometrics*, 34, 305–334.
- COSSLETT, S. R. (1981a): "Efficient estimation of discrete choice models," in *Structural analysis of discrete data with econometric applications*, ed. by C. F. Manski, and D. McFadden, pp. 51–111. MIT Press, Cambridge, MA.
- (1981b): "Maximum likelihood estimation for choice-based samples," *Econometrica*, 49, 1289–1316.
- (1991): "Efficient estimation from endogenously stratified samples with prior information on marginal probabilities," Manuscript, economics.sbs.ohio-state.edu/scosslett/papers/cbsample1.pdf.
- (1993): "Estimation from endogenously stratified samples," in *Handbook of Statistics*, ed. by G. Maddala, C. Rao, and H. Vinod, vol. 11, pp. 1–43. Elsevier Science, The Netherlands.
- CRAGG, J. G. (1983): "More efficient estimation in the presence of heteroscedasticity of unknown form," *Econometrica*, 49, 751–764.
- DEATON, A. (1997): *The analysis of household surveys. A microeconomic approach to development policy*. Johns Hopkins University Press.
- DEMETS, D., AND M. HALPERIN (1977): "Estimation of a simple regression coefficient in samples arising from a subsampling procedure," *Biometrics*, 33, 47–56.
- EL-BARMI, H., AND M. ROTHMANN (1998): "Nonparametric estimation in selection biased models in the presence of estimating equations," *Nonparametric Statistics*, 9, 381–399.
- GILL, R. D., Y. VARDI, AND J. A. WELLNER (1988): "Large sample theory of empirical distributions in biased sampling models," *Annals of Statistics*, 16, 1069–1112.
- HAUSMAN, J. A., AND D. A. WISE (1981): "Stratification on endogenous variables and estimation: The Gary income maintenance experiment," in *Structural analysis of discrete*

- data with econometric applications*, ed. by C. F. Manski, and D. McFadden, pp. 365–391. MIT Press, Cambridge, MA.
- HECKMAN, J. J. (1976): “The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models,” *Annals of Economic and Social Measurement*, 5, 475–492.
- (1979): “Sample selection bias as a specification error,” *Econometrica*, 47, 153–161.
- HIROSE, Y. (2007): “ M -estimators in semi-parametric multi-sample models,” Manuscript, sms.victoria.ac.nz/foswiki/pub/Main/ResearchReportSeries/mscs08-05.pdf.
- HIROSE, Y., AND A. J. LEE (2008): “Semi-parametric efficiency bounds for regression models under generalised case-control sampling: The profile likelihood approach,” *Annals of the Institute of Statistical Mathematics*, 62, 1023–1052.
- HOLT, D., T. M. F. SMITH, AND P. D. WINTER (1980): “Regression analysis of data from complex surveys,” *Journal of The Royal Statistical Society, Series A*, 143, 474–487.
- IMBENS, G. W. (1992): “An efficient method of moments estimator for discrete choice models with choice-based sampling,” *Econometrica*, 60, 1187–1214.
- IMBENS, G. W., AND T. LANCASTER (1996): “Efficient estimation and stratified sampling,” *Journal of Econometrics*, 74, 289–318.
- JEWELL, N. P. (1985): “Least squares regression with data arising from stratified samples of the dependent variable,” *Biometrika*, 72, 11–21.
- KALBFLEISCH, J. D., AND J. F. LAWLESS (1988): “Estimation of reliability in field-performance studies (with discussion),” *Technometrics*, 30, 365–388.
- KITAMURA, Y., G. TRIPATHI, AND H. AHN (2004): “Empirical likelihood based inference in conditional moment restriction models,” *Econometrica*, 72, 1667–1714.
- MANSKI, C. F. (1989): “Anatomy of the selection problem,” *Journal of Human Resources*, 24, 343–360.
- (1995): *Identification Problems in the Social Sciences*. Harvard University Press.
- MANSKI, C. F., AND S. R. LERMAN (1977): “The estimation of choice probabilities from choice based samples,” *Econometrica*, 45, 1977–1988.
- MANSKI, C. F., AND D. MCFADDEN (1981): “Alternative estimators and sample design for discrete choice analysis,” in *Structural analysis of discrete data with econometric applications*, ed. by C. F. Manski, and D. McFadden, pp. 2–50. MIT Press, Cambridge, MA.
- NEWBY, W. K. (1993): “Efficient estimation of models with conditional moment restrictions,” in *Handbook of Statistics, vol. 11*, ed. by G. S. Maddala, C. R. Rao, and H. D. Vinod, pp. 2111–2245. Elsevier Science B.V., The Netherlands.
- OWEN, A. B. (1988): “Empirical likelihood ratio confidence intervals for a single functional,” *Biometrika*, 75, 237–249.
- (2001): *Empirical likelihood*. Chapman and Hall/CRC.

- (2013): “Self-concordance for empirical likelihood,” *Canadian Journal of Statistics*, 41, 387–397.
- PAGAN, A., AND A. ULLAH (1999): *Nonparametric econometrics*. Cambridge University Press.
- PATIL, G. P., AND C. R. RAO (1977): “The weighted distributions: A survey of their applications,” in *Applications of Statistics*, ed. by P. R. Krishnaiah, pp. 383–405. North Holland.
- (1978): “Weighted distributions and size-biased sampling with applications to wildlife populations and human families,” *Biometrics*, 34, 179–189.
- QIN, J. (1993): “Empirical likelihood in biased sample problems,” *Annals of Statistics*, 21, 1182–1196.
- QUESENBERRY, C. P., AND N. P. JEWELL (1986): “Regression analysis based on stratified samples,” *Biometrika*, 73, 605–614.
- ROBINSON, P. M. (1987): “Asymptotically efficient estimation in the presence of heteroskedasticity of unknown form,” *Econometrica*, 55, 875–891.
- SCOTT, A. J., AND C. J. WILD (1986): “Fitting logistic models under case-control or choice based sampling,” *Journal of The Royal Statistical Society, Series B*, 48, 170–182.
- SEVERINI, T. A., AND G. TRIPATHI (2013): “Semiparametric efficiency bounds for microeconomic models: A survey,” *Foundations and Trends in Econometrics*, 6, 163–397.
- TRIPATHI, G. (1999): “A matrix extension of the Cauchy-Schwarz inequality,” *Economics Letters*, 63, 1–3.
- (2011a): “Generalized method of moments (GMM) based inference with stratified samples when the aggregate shares are known,” *Journal of Econometrics*, 165, 258–265.
- (2011b): “Moment based inference with stratified data,” *Econometric Theory*, 27, 47–73.
- TRIPATHI, G., AND Y. KITAMURA (2003): “Testing conditional moment restrictions,” *Annals of Statistics*, 31, 2059–2095.
- VARDI, Y. (1982): “Nonparametric estimation in the presence of length bias,” *Annals of Statistics*, 10, 616–620.
- (1985): “Empirical distributions in selection biased models,” *Annals of Statistics*, 13, 178–203.
- VARRETTE, S., P. BOUVRY, H. CARTIAUX, AND F. GEORGATOS (2014): “Management of an Academic HPC Cluster: The UL Experience,” in *Proceedings of the 2014 International Conference on High Performance Computing and Simulation (HPCS 2014)*, Bologna, Italy. IEEE.
- WOOLDRIDGE, J. M. (1999): “Asymptotic properties of weighted M-estimators for variable probability samples,” *Econometrica*, 67, 1385–1406.
- (2001): “Asymptotic properties of weighted M-estimators for standard stratified samples,” *Econometric Theory*, 17, 451–470.

——— (2010): *Econometric analysis of cross section and panel data*. MIT Press, Cambridge, Massachusetts, 2nd edn.

CENTER FOR RESEARCH IN ECONOMICS AND MANAGEMENT (CREA), FACULTY OF LAW, ECONOMICS AND FINANCE, UNIVERSITY OF LUXEMBOURG, L-1511, LUXEMBOURG.

E-mail address: `antonio.cosma@uni.lu`

CENTER FOR RESEARCH IN ECONOMICS AND MANAGEMENT (CREA), FACULTY OF LAW, ECONOMICS AND FINANCE, UNIVERSITY OF LUXEMBOURG, L-1511, LUXEMBOURG.

E-mail address: `andrei.kostyrka@uni.lu`

CENTER FOR RESEARCH IN ECONOMICS AND MANAGEMENT (CREA), FACULTY OF LAW, ECONOMICS AND FINANCE, UNIVERSITY OF LUXEMBOURG, L-1511, LUXEMBOURG.

E-mail address: `gautam.tripathi@uni.lu`