UNIVERSIDADE DE ÉVORA

DEPARTAMENTO DE ECONOMIA

DOCUMENTO DE TRABALHO Nº **2005/06**
April

Two-step Empirical Likelihood Estimation under Stratified Sampling when Aggregate Information is Available

Joaquim J. S. Ramalho*
Universidade de Évora, Departamento de Economia, CEMAPRE

Esmeralda A. Ramalho
Universidade de Évora, Departamento de Economia, CEMAPRE

This draft: April 2005

* The authors gratefully acknowledge partial financial support from Fundação para a Ciência e Tecnologia, program POCTI, partially funded by FEDER. Helpful comments and suggestions were given by seminar participants at the 7th Conference CEMAPRE, 2003, Lisbon. Address for correspondence: JoaquimJ. S. Ramalho, Universidade de Évora, Departmento de Economia, 7000-803 ÉVORA, Portugal. E-mail:jsr@uevora.pt.

UNIVERSIDADE DE ÉVORA
DEPARTAMENTO DE ECONOMIA
Largo dos Colegiais, 2 – 7000-803 Évora – Portugal
Tel.: +351 266 740 894 Fax: +351 266 742 494
www.decon.uevora.pt wp.economia@uevora.pt

Abstract:

Empirical likelihood (EL) is appropriate to estimate moment condition models when a random sample from the target population is available. However, many economic surveys are subject to some form of stratification, in which case direct application of EL will produce inconsistent estimators. In this paper we propose a two-step EL (TSEL) estimator to deal with stratified samples in models defined by unconditional moment restrictions in presence of some aggregate information, which may consist, for example, of the mean and the variance of the variable of interest and/or the explanatory variables.
A Monte Carlo simulation study reveals promising results for many versions of the TSEL estimator.

Palavras-chave/Keywords: Stratified Sampling, Empirical Likelihood, Weighted Estimation, Auxiliary Information

Classificação JEL/JEL Classification: C13, C51.

1 Introduction

In many research settings, economists often have to work with data which were not randomly drawn from the population they plan to study. Indeed, many economic datasets are the result of complex survey designs involving some form of stratification, rare alternatives being usually oversampled in order to reduce survey costs and improve the accuracy of econometric analysis; see *inter alia* Artis, Ayuso and Guillen (1999), Early (1999), Donkers, Franses and Verhoef (2003), and Kitamura, Yamamoto and Sakai (2003) for some interesting empirical cases where stratified data were collected. In this context, the available data provides a related but distorted picture of the features of the target population since the sampling distribution differs from the underlying distribution for which inference is to be drawn. Therefore, if the data are analyzed without regard to the sample design, seriously misleading inference results are, in general, obtained. Hence, unconventional estimation procedures are required to handle correctly data arising from stratified sampling.

For parametric and regression models, several estimation methods adequate to deal with stratified data have been proposed. The most well known estimators are likelihood-based, requiring the specification of the distribution of the variable of interest conditional on the covariates; see, for example, Manski and Lerman (1977), Manski and McFadden (1981), Cosslett (1981a,b), Imbens (1992), and Imbens and Lancaster (1996). For regression models, Holt, Smith and Winter (1980), Nathan and Holt (1980), and Hausman and Wise (1981) suggested weighted least squares approaches. Recently, Wooldridge (1999, 2001) proposed the use of weighted M-estimation, which is applicable to both classes of models, including some of the cited estimators as special cases. Naturally, the consequences of ignoring the sampling scheme extend to any other kind of econometric model. In this paper we focus on cases where the model to be estimated is defined by a set of unconditional moment restrictions, a setting only considered so far, to the best of our knowledge, by Tripathi (2004).

Tripathi (2004) considered empirical likelihood (EL) estimation of moment condition models, suggesting some adaptations to the standard EL methodology in order to take into account the sampling design. Those modifications ensure that the resultant EL estimators are consistent and efficient but require either the availability of an auxiliary dataset or knowledge on the marginal strata probabilities (the fraction of the population lying in each stratum). In this paper we assume that the former may be impossible or too expensive to produce and the latter are unknown, and propose EL estimators which merely require the availability of some other aggregate information on the target population. This auxiliary information may consist, for example, of the mean and the variance of the variable of interest and/or some covariates. We assume that this information is exact, since often it may be estimated very precisely from large survey data such as a census.

Building on Hellerstein and Imbens' (1999), we propose a two-step EL (TSEL) estimator. In the first step, we write the aggregate information available as moment restrictions and, using conventional EL estimation, calculate the set of weights that impose them in the sample. In the second step, we perform weighted EL estimation based on the weights estimated in the first step. In general, those weights do not allow the recovering of the target distribution, so the TSEL estimator is not consistent in most cases.[1] However, as the weighted stratified data used in the second step may be interpreted as a random sample drawn from an artificial population which shares the imposed moment restrictions with the target population, our estimator will tend to be less biased than conventional EL estimators. Obviously, the closer the target and the artificial populations are, the less the bias of the TSEL estimator is. Therefore, we investigate through a Monte Carlo simulation study which sets of commonly available aggregate information, if any, give rise to TSEL estimators displaying insignificant biases.

This paper is organized as follows. Section 2 briefly reviews the EL method and discusses why it fails to produce consistent estimators with stratified samples. Section 3 describes the TSEL estimator. Section 4 is dedicated to the Monte Carlo investigation. Section 5 concludes.

[1] An obvious exception occurs when the auxiliary information consists of the marginal strata probabilities.

2 Background

In this section we first describe the main characteristics of EL estimation of moment condition models when the sampling is random and then analyze some of the consequences of working with samples collected by either multinomial sampling (MS) or standard stratified sampling (SSS) in the moment condition framework.[2]

2.1 Empirical likelihood estimation of moment condition models

Consider a sample of $i = 1, ..., N$ individuals and let y be the variable of interest, continuous or discrete, and x a vector of k exogenous variables. Both y and x are random variables defined on $\mathcal{Y} \times \mathcal{X}$. Let $z \equiv (y, x)$, $\mathcal{Z} \equiv \mathcal{Y} \times \mathcal{X}$, and $f(z; \theta) \equiv dF(z; \theta)$ be the joint density of y and x, where θ is the k-vector of parameters of interest. In this paper we focus on models defined by a set of unconditional moment restrictions, i.e. models where all the information available about the population of interest can be summarized as

$$E_f [g(z, \theta_0)] = 0, \tag{1}$$

where $g(\cdot)$ is an m-dimensioned vector of unconditional moment indicators known up to θ, $m \geq k$, θ_0 is the unique solution of (1), and $E_f[\cdot]$ denotes expectation taken with respect to $f(z; \theta)$. A leading example in this area is instrumental variable estimation; see section 4 for a Monte Carlo simulation study involving these models. Throughout this paper we assume that the appropriateness of (1) is not in question.

With random sampling, despite $f(z; \theta)$ being unknown, there are several alternative methods that produce consistent estimators for θ_0, such as EL, where each observation is reweighted in such a way that all moment conditions are satisfied numerically in the sample. The EL estimator of θ_0 is obtained from maximization

[2]MS and SSS are two of the most well known stratified sampling schemes. There is another very popular sampling scheme, the so-called variable probability sampling. We do not consider it in this paper because we focus only on cases where the marginal strata probabilities are required for identification of the parameters of interest.

of the so-called log-EL function,

$$\log L(p_i) = \sum_{i=1}^{N} \log p_i, \tag{2}$$

subject to the set of restrictions

$$p_i \geq 0, \qquad \sum_{i=1}^{N} p_i = 1 \qquad \text{and} \qquad \sum_{i=1}^{N} p_i g(z_i, \theta) = 0, \tag{3}$$

where $p_i \equiv \Pr_f(z_i = z)$ is the probability mass assigned to each observed data point (z_i) by a discrete distribution with support on $\{z_1, ..., z_N\}$, and the last restriction is an empirical measure counterpart to the moment conditions (1). Under standard regularity conditions, optimization of (2) subject to (3) produces consistent estimators $\hat{\theta}$ and $\hat{F}_p(z) = \sum_{i=1}^{N} \hat{p}_i 1(z_i \leq z)$ for the vector of parameters of interest θ_0 and the distribution $F(z) = \Pr_f(z_i \leq z)$, respectively, where $1(\cdot)$ is the indicator function for the event $z_i \leq z$ and

$$\hat{p}_i = \frac{1}{N\left[1 + \hat{\lambda}' g\left(z_i, \hat{\theta}\right)\right]}, \tag{4}$$

where λ is the m-vector of Lagrange multipliers associated to the last restriction of (3). See Qin and Lawless (1994) and Imbens (1997) for more details, and Imbens (2002) for alternative computational procedures reliable for solving the optimization problem described by (2)-(3). In the Monte Carlo experiments undertaken in this paper we follow his penalty approach.

2.2 Stratification

From now on, assume that the population of interest is divided into J non-empty and possibly overlapping strata. Each stratum is designated as $\mathcal{C}_s = \mathcal{Y}_s \times \mathcal{X}_s$, with $s \in \mathcal{S}$, $\mathcal{S} = \{1, ..., J\}$, and $\mathcal{Y}_s$ and $\mathcal{X}_s$ defined as the subsets of, respectively, $\mathcal{Y}$ and $\mathcal{X}$ for which the observation (y, x) lies in $\mathcal{C}_s$. The proportion of the stratum s in the population is given by $Q_s = \int_{\mathcal{X}_s} \int_{\mathcal{Y}_s} f(z; \theta) \, dydx$.

In this paper we provide an unified approach for both MS and SSS by writing the density induced by the two sampling mechanisms as

$$f^*(z,s;\theta) = \frac{1}{\omega} f(z;\theta), \tag{5}$$

where, although we do not make it notationally explicit, $f^*(z,s;\theta)$ may be conditional on some variables and $\omega \equiv \omega(z,s) > 0$ is the weight that, multiplied by the sampled distribution, allows the recovering of the target distribution. In the MS scheme, considered *inter alia* by Manski and Lerman (1977), Manski and McFadden (1981), and Imbens (1992), it is assumed that the stratum indicators s are independently drawn from a multinomial distribution. The sampling agent randomly selects a stratum $\mathcal{C}_s$ with a pre-defined probability H_s, $0 < H_s < 1$, and, then, randomly samples from that stratum. Thus, for each sampling unit, the variable of interest, the covariates, and the stratum indicator are observed according to (5) with $\omega = Q_s/H_s$. With regard to the SSS scheme, the sampling agent fixes the number of observations N_s, $N_s > 0$, to be randomly collected from each stratum $\mathcal{C}_s$; see, for example, Cosslett (1981a,b), Hsieh, Manski and McFadden (1985), and Wooldridge (2001). In this sampling mechanism, as N_s is fixed by design, the sampling proportion of each strata, N_s/N, is also fixed. As Imbens and Lancaster (1996) showed, in this case inference may be based on the likelihood $f^*(z,s|N_s;\theta)$ which is given by (5) with $\omega = NQ_s/N_s$.

When the available dataset was collected according to MS or SSS, the population moment conditions (1) do not hold in the sample, even asymptotically, since

$$\begin{aligned} E_{f^*}[g(z,\theta_0)] &= \int_{\mathcal{Z}} g(z,\theta_0) f^*(z,s;\theta_0)\,dz \\ &= \int_{\mathcal{Z}} g(z,\theta_0) \frac{1}{\omega} f(z;\theta_0)\,dz \\ &= E_f\left[\frac{1}{\omega} g(z,\theta_0)\right], \end{aligned} \tag{6}$$

which is, in general, different from zero. Hence, imposing (1) in the sample would lead to inconsistent estimates of the parameters of interest, so direct application of EL is not possible with stratification. Note that this conclusion is valid both in presence of endogenous (sampling scheme where the dependent variable is among the design variables) and exogenous stratification (where the strata are designed only in terms of one or more covariates). Hence, in models defined by a set of unconditional moment restrictions, both types of stratification must be taken into account, unlike what happens in conditional structural models, the focus of most of the literature on stratified samples, where exogenous stratification is innocuous.

Most of the literature on stratified samples assumes that the weights ω are known, which implies that the marginal strata probabilities Q_s must also be known. In such case, it is straightforward to reconstruct the structure of the target population and, therefore, the use of weighted estimators arises as a natural approach to deal with stratified samples; see, for example, the pioneering work by Manski and Lerman (1977), where estimation is based on weighted scores, the weighted least squares approach by Nathan and Holt (1980) and others, and the recent proposal by Wooldridge (1999, 2001) on weighted M-estimation. Similarly, for moment condition models, Tripathi (2004) suggested a weighted EL (WEL) estimator, which is obtained by solving a program similar to that described by (2) and (3) but based on the reweighted moment indicators $g^*(z,\theta) = \omega g(z,\theta)$, since $E_{f^*}[g^*(z,\theta_0)] = E_f[g(z,\theta_0)] = 0$, see (1) and (6).

In case Q_s is not known, in parametric models it is possible to estimate it simultaneously with the parameters of interest and, thus, still perform weighted estimation; see *inter alia* Cosslett (1981a,b), Imbens (1992), and Imbens and Lancaster (1996). However, such approach is not possible in semi-parametric models like those analyzed in this paper, unless an auxiliary data set is available. Below we suggest a TSEL estimator which merely requires the availability of some aggregate information on the population of interest (other than Q_s).

3 Two-step empirical likelihood estimation

There are a few papers on the utilization of auxiliary information in econometrics but most of them focus only on efficiency issues. To the best of our knowledge, only Hellerstein and Imbens (1999) dealt with cases where the target and the sampled distributions differ. In contrast to our work, they considered only just-identified models estimated by least squares. Here, we extend their two-step procedure to the overidentified moment condition framework. In the first step, similarly to their approach, a set of weights that impose the aggregate information in the sample is calculated using standard EL estimation. In the second step, weighted EL estimation based on the weights estimated in the first step is performed.

3.1 First step: using the auxiliary information to estimate the weights

Consider an m_h-vector $\bar{h}(z)$ and assume that the analyst has exact knowledge of its expectation $h^* = E_f\left[\bar{h}(z)\right]$. Using this aggregate information, he or she may construct the set of moment indicators

$$h(z) = \bar{h}(z) - h^* \tag{7}$$

whose expectation taken under the target population is known to be zero,

$$E_f[h(z)] = 0. \tag{8}$$

For example, if the analyst has information about the mean of y, the noncentered covariance between y and x_1, where x_1 is a given covariate, and the probability of z being in a given subset $\mathcal{C}$ of the sample space, $h(z) = [y - E(y),\ yx_1 - E(yx_1),\ 1(z \in \mathcal{C}) - P(z \in \mathcal{C})]'$.

In our framework, due to stratification, $h(z)$ does not have, in general, expectation zero in the sampled population, $E_{f^*}[h(z)] \neq 0$. However, using conventional EL estimation, we may estimate a weighting scheme that imposes (8) in the sample. Thus, the first stage of our TSEL method corresponds to the optimization of the Lagrangian function

$$\mathcal{L}(\upsilon_i, \tau, \phi) = \sum_{i=1}^{N} \log \upsilon_i - \tau\left(\sum_{i=1}^{N} \upsilon_i - N\right) - \phi' \sum_{i=1}^{N} \upsilon_i h(z_i), \tag{9}$$

where τ and the m_h-vector ϕ are Lagrange multipliers and υ_i denotes the weights assigned to each observation. It is straightforward to show that this constrained optimization is equivalent to maximization of the unconstrained objective function

$$R(\phi) = \sum_{i=1}^{N} \log\left[1 + \phi' h(z_i)\right] \tag{10}$$

and that the estimated weights are given by

$$\tilde{\upsilon}_i = \frac{1}{1 + \tilde{\phi}' h(z_i)}, \tag{11}$$

where $\tilde{\phi}$ minimizes (10), solving

$$\sum_{i=1}^{N} \frac{h(z_i)}{1 + \tilde{\phi}' h(z_i)} = 0. \tag{12}$$

3.2 Second step: weighted EL estimation

The second stage of our TSEL method corresponds to weighted EL estimation. Due to stratification, the log-EL criterion function (2) must be redefined as

$$\log L(p_i^*) = \sum_{i=1}^{N} \log p_i^*, \tag{13}$$

where $p_i^* \equiv \Pr_{f^*}(z_i = z)$. Corresponding to the set of restrictions (3), we now impose $p_i^* \geq 0$ and $\sum_{i=1}^{N} p_i^* = 1$ but cannot impose $\sum_{i=1}^{N} p_i^* g(z_i, \theta) = 0$, since the moment conditions (1) do not hold in the observed data, only in the target population. However, when the estimated weights $\tilde{\upsilon}$ are sufficiently close the true weights ω, $E_{f^*}\left[g^*\left(z, \theta_0, \tilde{\phi}\right)\right] \simeq E_{f^*}\left[g^*(z, \theta_0)\right] = 0$, where $g^*\left(z, \theta, \tilde{\phi}\right) \equiv \tilde{\upsilon} g(z, \theta)$. Hence we may express the optimization problem (2)-(3) in terms of the available data by replacing (2) by (13) and the last restriction of (3) by $\sum_{i=1}^{N} p_i g^*\left(z_i, \theta, \tilde{\phi}\right)$. The new Lagrangian function is thus

$$\mathcal{L}(p_i^*, \gamma, \lambda, \theta) = \sum_{i=1}^{N} \log p_i^* - \gamma\left(\sum_{i=1}^{N} p_i^* - 1\right) - N\lambda' \sum_{i=1}^{N} p_i^* g^*\left(z_i, \theta, \tilde{\phi}\right). \tag{14}$$

The only practical difference between this and the Lagrangian function corresponding to the optimization problem (2)-(3) is that now the analysis is based on the weighted moment indicators $g^*\left(z, \theta, \tilde{\phi}\right)$, instead of $g(z, \theta)$.

Similarly to the random sampling case, see (4), solving the system of first-order conditions resultant from optimization of (14) we find

$$\hat{p}_i^* = \frac{1}{N\left[1+\hat{\lambda}' g^*\left(z_i, \theta, \tilde{\phi}\right)\right]}. \tag{15}$$

Concentrating out p_i^* from $\sum_{i=1}^N \log p_i^*$ and dropping irrelevant terms, we obtain the unconstrained EL criterion function

$$R(\theta, \lambda) = \sum_{i=1}^N \log\left[1+\lambda' g^*\left(z_i, \theta, \tilde{\phi}\right)\right], \tag{16}$$

optimization of which produces the EL estimators $\hat{\theta}$ and $\hat{\lambda}$ which satisfy the first-order conditions

$$\sum_{i=1}^N \frac{1}{1+\hat{\lambda}' \hat{g}^*\left(z_i, \hat{\theta}\right)} \left[\begin{array}{c} g^*\left(z_i, \hat{\theta}, \tilde{\phi}\right) \\ G^*\left(z_i, \hat{\theta}, \tilde{\phi}\right)' \hat{\lambda} \end{array}\right] = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tag{17}$$

where $G^*\left(z_i, \hat{\theta}, \tilde{\phi}\right) \equiv \frac{\partial g^*\left(z_i, \hat{\theta}, \tilde{\phi}\right)}{\partial \theta'}$, $i = 1, ..., N$.

3.3 Asymptotics

In general, $\tilde{\upsilon}_i \neq \omega_i$, even asymptotically, since the information provided by (8) is not enough to recover the structure of the target distribution. Therefore, in most cases, the TSEL estimator will not converge asymptotically to θ_0. Instead, it will be consistent for θ_a, the probability limit of the unweighted EL estimator of θ_0 in (1) based on a random sample from an artificial population with probability density function $f^a(z, s) = \frac{f^*(z, s; \theta)}{1+\phi_a' h(z)}$, where $\phi_a \equiv \arg\min_\phi E_{f^*}[R(\phi)]$. This artificial distribution may be seen as the distribution closest to the sampling distribution, using the EL metric, which satisfies the population moment conditions (8), that is $E_{f^a}[h(z)] = E_f[h(z)] = 0$. On the other hand, as the weights υ shift the sampling distribution towards the target distribution, the TSEL estimator will display, in general, less bias than conventional EL estimators.

Naturally, the ability of the TSEL estimator to reduce the bias of the EL estimator depends on the quantity and quality of the information contained in $h(z)$. The worst scenario occurs when the information supplied by $h(z)$ is irrelevant, which happens when $E_{f^*}[h(z)] = 0$, that is the sampling scheme was such that $h(z)$ has expectation zero both in the sample and in the target population. Indeed, in this case $\phi_a = 0$ and $f^a(z,s) = f^*(z,s;\theta)$, so two-step and conventional EL estimators will display the same bias. Conversely, the inconsistency of the TSEL estimator is completely eliminated when the auxiliary information consists of the marginal strata probabilities Q_s, that is $h(z) = 1(z \in \mathcal{C}_s) - Q_s$, $s = 1, ..., J-1$, which implies $\tilde{\upsilon}_i = \omega_i$, $i = 1, ..., n$, and, hence, $f^a(z,s) = f(z,s;\theta)$. In such case, TSEL is equivalent to Tripathi's (2004) WEL estimation, producing identical results.

As shown in the appendix, for overidentified problems we have

$$\sqrt{N}\left(\hat{\theta} - \theta_a\right) \xrightarrow{d} \mathcal{N}\left[0, \left(G^{*\prime}V_{gg}^{*-1}G^*\right)^{-1}\left[I - G^{*\prime}V_{gg}^{*-1}V_{gh}^*V_{hh}^{*-1}V_{hg}^*V_{gg}^{*-1}G^*\left(G^{*\prime}V_{gg}^{*-1}G^*\right)^{-1}\right]\right], \tag{18}$$

where

$$\begin{aligned}\begin{bmatrix} V_{gg}^* & V_{gh}^* \\ V_{hg}^* & V_{hh}^* \end{bmatrix} &= E_{f^*}\begin{bmatrix} g^*(z_i,\theta_a,\phi_a)\, g^*(z_i,\theta_a,\phi_a)' & g^*(z_i,\theta_a,\phi_a)\, h^*(z_i,\phi_a) \\ h^*(z_i,\phi_a)\, g^*(z_i,\theta_a,\phi_a) & h^*(z_i,\phi_a) \end{bmatrix}, \\ G^* &= E_{f^*}\left[\frac{g^*(z_i,\theta_a,\phi_a)}{\partial\theta'}\right], \text{ and } h^*(z_i,\phi_a) = \upsilon h(z_i).\end{aligned}$$

In case $s = k$, it is straightforward to show that the above result may be simplified to

$$\sqrt{N}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} \mathcal{N}\left[0, G^{*-1}\left(V_{gg}^* - V_{gh}^*V_{hh}^{*-1}V_{hg}^*\right)G^{*-1}\right],$$

which, specialized to least squares, corresponds to the asymptotic distribution derived by Hellerstein and Imbens (1999).

4 Monte Carlo Simulation Study

4.1 Experimental Design

In this section we conduct a Monte Carlo simulation study in order to examine the finite sample properties of several versions of the TSEL estimator. We considered a linear instrumental variable model described by the moment conditions

$$E_f \left[W' \left(y - X\theta_0 \right) \right] = 0, \tag{19}$$

where W is an $(N \times 4)$ matrix of instruments and y and X are N-vectors of observations on a dependent and an explanatory variable, respectively. We generate these variables from

$$W_j = \gamma_j \epsilon + v_j, \; j = 1, ..., 4, \tag{20}$$

$$X = \lambda u + \epsilon \tag{21}$$

and

$$y = X\theta_0 + u, \tag{22}$$

where ϵ, v_j and u are random disturbances independently generated from a $\mathcal{N}(0,1)$ distribution, $\theta_0 = 1$, and λ and γ_j are fixed parameters whose values were chosen in such a way that the correlations ρ_{xu} between X and u and ρ_{xw_j} between X and W_j are both equal to 0.5.[3]

Using this setting, we generated randomly a "super-population" of size $15N$, which can be interpreted as the target population. Then, various sampling designs were used to select stratified samples of size $N = 300$. All of them were based upon a partition of the population into four strata, considering alternately as design variable y or X. In the former case the strata were defined as $\mathcal{C}_1 = (-\infty, -3.072) \times \mathcal{X}$, $\mathcal{C}_2 = (-3.072, 0) \times \mathcal{X}$, $\mathcal{C}_3 = (0, 3.072) \times \mathcal{X}$ and $\mathcal{C}_4 = (3.072, \infty) \times \mathcal{X}$. In the other, we considered $\mathcal{C}_1 = \mathcal{Y} \times (-\infty, -1.899)$, $\mathcal{C}_2 = \mathcal{Y} \times (-1.899, 0)$, $\mathcal{C}_3 = \mathcal{Y} \times (0, 1.899)$, $\mathcal{C}_4 = \mathcal{Y} \times (1.899, \infty)$. In both cases the limits of each strata were chosen in order to

[3] Note that $\lambda = \frac{\rho_{xu}}{\sqrt{1-\rho_{xu}^2}}$ and $\gamma_j = \rho_{xw_j} \sqrt{\frac{1+\lambda^2}{1-(1+\lambda^2)\rho_{xw_j}^2}}$.

produce $Q = (0.05, 0.45, 0.45, 0.05)$. For both classes of stratification we examined the effects of two alternative allocations of the total sample over the four strata: equal shares allocation, $H = (0.25, 0.25, 0.25, 0.25)$, and U-shaped allocation, $H = (0.4, 0.1, 0.1, 0.4)$. For each experiment we generated 1000 independent Monte Carlo samples.

For each experimental design, we computed four different TSEL estimators, labelled TSELa, TSELb, TSELc, and TSELd, which assume knowledge on, respectively, 1, 2, 3, or 4 moments of y (in case of endogenous stratification) or X (exogenous stratification). For example, the TSELd estimator is based on

$$h(y) = \begin{bmatrix} y \\ y^2 - 3.488 \\ y^3 \\ y^4 - 36.499 \end{bmatrix} \text{ or } h(X) = \begin{bmatrix} X \\ X^2 - 1.333 \\ X^3 \\ X^4 - 5.333 \end{bmatrix}. \quad (23)$$

In each case we computed also the EL and WEL estimators, which will act as benchmarks for our estimators. Note that the latter estimator may also be obtained as a TSEL estimator based on

$$h(y) = \begin{bmatrix} 1(y \in \mathcal{C}_1) - 0.05 \\ 1(y \in \mathcal{C}_2) - 0.45 \\ 1(y \in \mathcal{C}_3) - 0.45 \end{bmatrix} \text{ or } h(X) = \begin{bmatrix} 1(X \in \mathcal{C}_1) - 0.05 \\ 1(X \in \mathcal{C}_2) - 0.45 \\ 1(X \in \mathcal{C}_3) - 0.45 \end{bmatrix}. \quad (24)$$

4.2 Results

Table 1 reports for each estimator the mean and median bias, standard error (SE), root mean squared error (RMSE), median absolute error (MAE) and the 0.05 and 0.95 quantiles of its Monte Carlo distribution. Clearly, in all cases, if the sampling design is not taken in account, the resultant estimators are substantially biased. As can be seen and we stressed before, in the setting considered in this paper, unlike what happens in conditional models, stratification on X, if not accounted for, gives rise to seriously misleading results.

Table 1 about here

As expected, the WEL estimator is approximately unbiased in all cases. With regard to the TSEL estimators, when only the first moment is known, there is no improvement over conventional EL estimation. This is not surprising since $E_{f^*}(y) = E_f(y)$ (or $E_{f^*}(X) = E_f(X)$) and, hence, the results obtained are virtually identical to standard EL estimation. In contrast, the other three TSEL estimators performed in a very promising way, displaying less SE, RMSE, and MAE than the WEL estimator in almost all cases. In terms of bias, only the TSELd estimator appears to be also approximately unbiased in all experiments, although knowledge of only the first two moments allows a very substantial reduction on the bias of the EL estimator. In fact, the availability of information on the fourth moment is crucial to obtain unbiased estimators only in the U-shaped allocation sampling design. With an equal shares allocation, information on the first two moments seems to be sufficient to capture the main characteristics of the population structure and, thus, to obtain unbiased estimators.

5 Conclusion

In this paper we proposed TSEL estimators to deal with stratified samples in the EL framework when the marginal strata probabilities are unknown but other aggregate information on some features of the population of interest is available. The ability of these estimators to reduce the bias of the standard EL estimator depends crucially on the relevance of the auxiliary information available, as illustrated by the Monte Carlo study. Most versions of the TSEL estimator produced very satisfactory bias results and performed better in terms of SE, RMSE, and MAE than Tripathi's (2004) WEL estimator, which requires knowledge on the marginal strata probabilities.

Some extensions to this paper are straightforward. First, the assumption that the auxiliary population information is exactly known can be dropped with minor adaptations. Indeed, if that information was estimated from an auxiliary random sample not very large, the sampling error incurred in their estimation could be accounted for using similar procedures to those proposed by Imbens and Lancaster (1994), and Hellerstein and Imbens (1999). Second, in this paper we have focussed on EL estimation. However, any other of Newey and Smith's (2004) generalized EL estimators could be adapted in a similar way.

References

Artis, M., Ayuso, M. and Guillén, M. (1999). 'Modelling Different Types of Automobile Insurance Fraud Behaviour in the Spanish Market', *Insurance: Mathematics and Economics*, Vol. 24, pp. 67-81.

Cosslett, S. (1981a). 'Efficient Estimation of Discrete-Choice Models', in: Manski, C. and McFadden, D. (eds.), *Structural Analysis of Discrete Data with Econometric Applications*, The MIT Press, pp. 51-111.

Cosslett, S. (1981b). 'Maximum Likelihood Estimator for Choice-Based Samples', *Econometrica*, Vol. 49, No. 5, pp. 1289-1316.

Donkers, B., Franses, P. H. and Verhoef, P. C. (2003). 'Selective Sampling for Binary Choice Models', *Journal of Marketing Research*, Vol. 40, No. 4, pp. 492-497.

Early, D. W. (1999). 'A Microeconomic Analysis of Homelessness: An Empirical Investigation Using Choice-Based Sampling', *Journal of Housing Economics*, Vol. 8, No. 4, pp. 312-327.

Hausman, J. and Wise, D. (1981). 'Stratification on Endogenous Variables and Estimation: The Gary Income Maintenance Experiment', in: Manski, C. and McFadden, D. (eds.), *Structural Analysis of Discrete Data with Econometric Applications*, The MIT Press, pp. 365-391.

Hellerstein, J. K. and Imbens, G. W. (1999). 'Imposing Moment Restrictions from Auxiliary Data by Weighting', *Review of Economics and Statistics*, Vol. 81, No. 1, pp. 1-14.

Holt, D., Smith, T. M. F. and Winter, P. D. (1980). 'Regression Analysis of Data from Complex Surveys', *Journal of the Royal Statistical Society*, Series A, Vol. 143, No. 4, pp. 474-487.

Hsieh, D. A., Manski, C. F. and McFadden, D. (1985). 'Estimation of Response Probabilities from Augmented Retrospective Observations', *Journal of the American Statistical Association*, Vol. 80, No. 391, pp. 651-662.

Imbens, G. (1992). 'An Efficient Method of Moments Estimator for Discrete Choice Models with Choice-Based Sampling', *Econometrica*, Vol. 60, No. 5, pp. 1187-1214.

Imbens, G. W. (1997). 'One-Step Estimators for Over-Identified Generalised Method of Moments Models', *Review of Economic Studies*, Vol. 64, No. 3, pp. 359-383.

Imbens, G. W. (2002). 'Generalized Method of Moments and Empirical Likelihood', *Journal of Business & Economic Statistics*, Vol. 20, No. 4, pp. 493-506.

Imbens, G. W. and Lancaster, T. (1994). 'Combining Micro and Macro Data in Microeconometric Models', *Review of Economic Studies*, Vol. 61, No. 4, pp. 655-680.

Imbens, G. W. and Lancaster, T. (1996). 'Efficient Estimation and Stratified Sampling', *Journal of Econometrics*, Vol. 74, No. 2, pp. 289-318.

Kitamura, R., Yamamoto, T. and Sakai, H. (2003). 'A Methodology for Weighting Observations from Complex Endogenous Sampling', *Transportation Research*, Part B, Vol. 37, No. 4, pp. 387-401.

Manski, C. and Lerman, S. (1977). 'The Estimation of Choice Probabilities from Choice Based Samples', *Econometrica*, Vol. 45, No. 8, pp. 1977-1988.

Manski, C. and McFadden, D. (1981). 'Alternative Estimators and Sample Designs for Discrete Choice Analysis', in Manski, C. and McFadden, D. (eds.), *Structural Analysis of Discrete Data with Econometric Applications*, The MIT Press, pp. 2-50.

Nathan, G. and Holt, D. (1980). 'The Effect of Survey Design on Regression Analysis', *Journal of the Royal Statistical Society*, Series B, Vol. 42, No. 3, pp. 377-386.

Newey, W. K. and Smith, R. J. (2004). 'Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators', *Econometrica*, Vol. 72, No. 1, pp. 219-255.

Qin, J. and Lawless, J. (1994). 'Empirical Likelihood and General Estimating Equations', *Annals of Statistics*, Vol. 22, No. 1, pp. 300-325.

Tripathi, G. (2004). 'Moment Based Inference with Stratified Data', Mimeo.

Tripathi, G. and Kitamura, Y. (2003). 'Testing Conditional Moment Restrictions', *Annals of Statistics*, Vol. 31, No. 6, pp. 2059-2095.

Wooldridge, J. M. (1999). 'Asymptotic Properties of Weighted M-Estimators for Variable Probability Samples', *Econometrica*, Vol. 67, No. 6, pp. 1385-1406.

Wooldridge, J. M. (2001). 'Asymptotic Properties of Weighted M-Estimators for Standard Stratified Samples', *Econometric Theory*, Vol. 17, No. 2, pp. 451-470.

6 Appendix: asymptotic distribution of the TSEL estimator

Define $h_i^*(\phi) = \frac{h(z_i)}{1+\phi' h(z_i)}$, $g_i^*(\theta, \phi) = \frac{g(z_i,\theta)}{1+\phi' h(z_i)}$, $G_i^*(\theta, \phi) = \frac{\partial g_i^*(\theta,\phi)}{\partial \theta}$, $\bar{g}_i(\lambda, \theta, \phi) = \frac{g_i^*(\theta,\phi)}{1+\lambda' g_i^*(\theta,\phi)}$, and $\bar{G}_i(\lambda, \theta, \phi) = \frac{G_i^*(\theta,\phi)}{1+\lambda' g_i^*(\theta,\phi)}$. The TSEL estimator can be characterized as the solution to the system of equations

$$\psi\left(\hat{\lambda}, \hat{\theta}, \hat{\phi}\right) = \sum_{i=1}^{N} \begin{bmatrix} \bar{g}_i\left(\hat{\lambda}, \hat{\theta}, \hat{\phi}\right) \\ \bar{G}_i\left(\hat{\lambda}, \hat{\theta}, \hat{\phi}\right)' \hat{\lambda} \\ h_i^*\left(\hat{\phi}\right) \end{bmatrix} = 0,$$

where the two first equations define the estimators $\left(\hat{\lambda}, \hat{\theta}\right)$ obtained in the second step, see (17), and the last equation defines the Lagrange multiplier $\hat{\phi}$ calculated in the first step, see (12).

Expanding the set of moment indicators around the probability limits of $\left(\hat{\lambda}, \hat{\theta}, \hat{\phi}\right)$, $(0, \theta_a, \phi_a)$, yields

$$
\begin{aligned}
0 &= \frac{1}{N}\begin{bmatrix} \sum g_i^*(\theta_a, \phi_a) \\ 0 \\ \sum h_i^*(\phi_a) \end{bmatrix} + \\
&\quad \frac{1}{N}\begin{bmatrix} -\sum g_i^*(\theta_a, \phi_a)\, g_i^*(\theta_a, \phi_a)' & \sum G_i^*(\theta_a, \phi_a) & -\sum g_i^*(\theta_a, \phi_a)\, h_i^*(\phi_a)' \\ \sum G_i^*(\theta_a, \phi_a)' & 0 & 0 \\ 0 & 0 & -\sum h_i^*(\phi_a)\, h_i^*(\phi_a)' \end{bmatrix} \begin{bmatrix} \hat{\lambda} \\ \hat{\theta} - \theta_a \\ \hat{\phi} - \phi_a \end{bmatrix}.
\end{aligned}
$$

Thus:

$$
\begin{aligned}
\sqrt{N}\begin{bmatrix} \hat{\lambda} \\ \hat{\theta} - \theta_a \\ \hat{\phi} - \phi_a \end{bmatrix} &= -\begin{bmatrix} -V_{gg}^* & G^* & -V_{gh}^* \\ G^{*\prime} & 0 & 0 \\ 0 & 0 & -V_{hh}^* \end{bmatrix}^{-1} \sqrt{N}\begin{bmatrix} g_N^*(\theta_a, \phi_a) \\ 0 \\ h_N^*(\phi_a) \end{bmatrix} \\
&= -\begin{bmatrix} -V_{gg}^{*-1} + V_{gg}^{*-1}G^*\left(G^{*\prime}V_{gg}^{*-1}G^*\right)^{-1}G^{*\prime}V_{gg}^{*-1} & V_{gg}^{*-1}G^*\left(G^{*\prime}V_{gg}^{*-1}G^*\right)^{-1} & V_{gg}^{*-1}V_{gh}^*V_{hh}^{*-1} - V_{gg}^{*-1}G^*\left(G^{*\prime}V_{gg}^{*-1}G^*\right)^{-1}G^{*\prime}V_{gg}^{*-1}V_{gh}^*V_{hh}^{*-1} \\ \left(G^{*\prime}V_{gg}^{*-1}G^*\right)^{-1}G^{*\prime}V_{gg}^{*-1} & \left(G^{*\prime}V_{gg}^{*-1}G^*\right)^{-1} & -\left(G^{*\prime}V_{gg}^{*-1}G^*\right)^{-1}G^{*\prime}V_{gg}^{*-1}V_{gh}^*V_{hh}^{*-1} \\ 0 & 0 & -V_{hh}^{*-1} \end{bmatrix} \\
&\quad \sqrt{N}\begin{bmatrix} g_N^*(\theta_a, \phi_a) \\ 0 \\ h_N^*(\phi_a) \end{bmatrix}.
\end{aligned}
$$

Result (18) follows straightforwardly.

Table 1: Monte Carlo results

Design variable	Estimator	Bias		Quantiles		SE	RMSE	MAE
		Mean	Median	0.05	0.95			
y		H = {0.25, 0.25, 0.25, 0.25}						
	EL	.375	.374	1.312	1.436	.038	.377	.374
	TSELa	.375	.375	1.312	1.436	.038	.376	.375
	TSELb	.004	.005	0.901	1.106	.065	.065	.044
	TSELc	.004	.005	0.901	1.106	.065	.065	.044
	TSELd	.003	.004	0.896	1.103	.065	.065	.045
	WEL	.000	.002	0.868	1.119	.077	.077	.053
		H = {0.4, 0.1, 0.4, 0.1}						
	EL	.488	.488	1.439	1.536	.030	.489	.488
	TSELa	.488	.488	1.439	1.536	.029	.489	.488
	TSELb	.032	.034	0.841	1.220	.113	.118	.078
	TSELc	.032	.034	0.842	1.219	.114	.118	.079
	TSELd	.011	.011	0.840	1.176	.104	.104	.071
	WEL	.004	.012	0.791	1.194	.123	.124	.082
X		H = {0.25, 0.25, 0.25, 0.25}						
	EL	.262	.264	1.207	1.314	.032	.264	.264
	TSELa	.262	.264	1.207	1.314	.032	.264	.264
	TSELb	.003	.004	0.880	1.123	.072	.072	.045
	TSELc	.003	.005	0.881	1.121	.072	.072	.045
	TSELd	.002	.003	0.881	1.120	.071	.071	.045
	WEL	.000	.005	0.866	1.129	.079	.079	.050
		H = {0.4, 0.1, 0.4, 0.1}						
	EL	.314	.313	1.274	1.355	.026	.315	.313
	TSELa	.314	.313	1.273	1.355	.026	.315	.313
	TSELb	.024	.035	0.793	1.207	.129	.131	.091
	TSELc	.024	.035	0.794	1.207	.128	.131	.090
	TSELd	.008	.011	0.813	1.182	.112	.112	.077
	WEL	-.003	.001	0.788	1.199	.127	.127	.084

Notes: results based on 1000 Monte Carlo replications, $N = 300$, and $\theta_0 = 1$.