Inference in Conditional Moment Restriction Models When There Is Selection due to Stratification

(forthcoming in *Advances in Econometrics*, vol. 39)

Antonio Cosma, Andreï V. Kostyrka, Gautam Tripathi

UNIVERSITÉ DU LUXEMBOURG
Centre for Research in Economics
and Management (CREA)

DSEF Ph. D. workshop, September 25, 2018

- Construct **efficient estimators** in models defined by **conditional moment restrictions** under **variable probability (VP) sampling**
- Identification and estimation of models with conditional moment restrictions by Smooth Empirical Likelihood (SEL)
- Inference
- Example: linear regression model under VP sampling

Variable Probability (VP) Sampling

- Often, the data economists plan to use are not drawn from the population of interest, but a closely related one.
- VP sampling is used in telephone surveys, or oversampling of specific categories to improve precision of estimates (e. g. high- vs low-income households).
- Other sampling schemes:
    - Standard Stratification (SS)
    - Multinomial Sampling (MNS)

Some Notation

- *Target population*, i. e. the population of interest:
  - $Z^* = (Y^*, X^*)$ is a random vector, $Z^* \sim P^*$
  - $\mathbb{C}_1, \ldots, \mathbb{C}_L$ is a partition of the support of Z^* ($\operatorname{supp} Z^*$)
- *Realised population*, i. e. the data actually collected:
  - Each draw is retained with probability p_l according to the stratum $\mathbb{C}_l$ to which it belongs
  - The *retained* random vector $Z = (Y, X)$ follows the law P:

$$P(Z \in B) = \sum_{l=1}^{L} \frac{p_l}{b^*} \int_B \mathbb{I}_{\mathbb{C}_l}(z) \, \mathrm{d}P^*(z),$$

  where $b^* \stackrel{\text{def}}{=} \sum_l p_l Q_l^*$ and $Q_l^* \stackrel{\text{def}}{=} P^*(Z^* \in \mathbb{C}_l) > 0$.
- Under VP sampling, the support of the distribution of the realised population **is the same** as the support of the target population.

Exogenous vs Endogenous Stratification

Let Y^* be an endogenous variable and X^* an exogenous variable, in the target population.

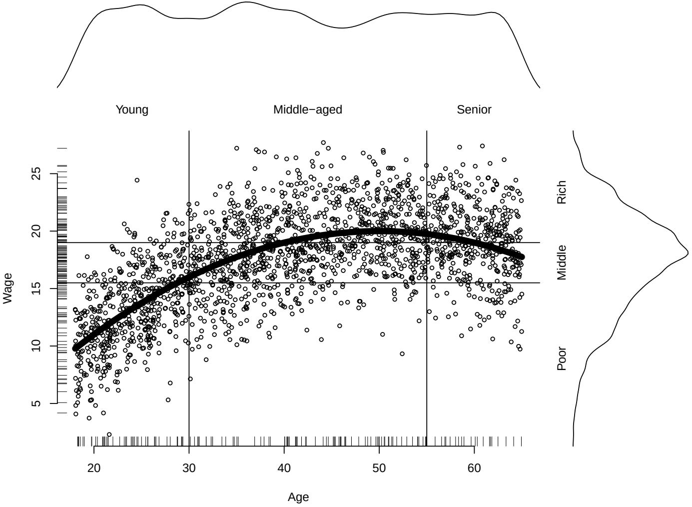
$$\operatorname{supp} Y^* = \bigcup_j \mathbb{A}_j, \quad \operatorname{supp} X^* = \bigcup_m \mathbb{B}_m.$$

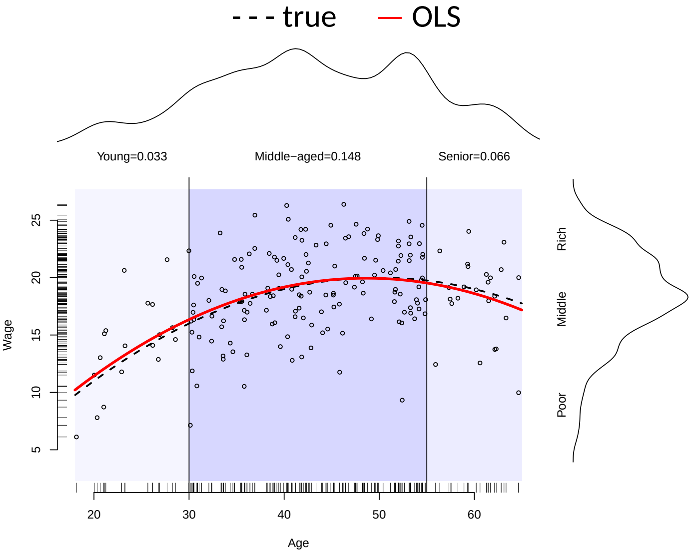
Exogenous and *endogenous* stratification are special cases of:

$$\operatorname{supp}(Y^*, X^*) = \begin{cases} \bigcup_j \bigcup_m (\mathbb{A}_j \times \mathbb{B}_m) & \text{if both } Y^* \text{ and } X^* \text{ are stratified,} \\ \bigcup_j (\mathbb{A}_j \times \operatorname{supp} X^*) & \text{if only } Y^* \text{ is stratified: } \textit{endogenous}, \\ \bigcup_m (\operatorname{supp} Y^* \times \mathbb{B}_m) & \text{if only } X^* \text{ is stratified: } \textit{exogenous}. \end{cases}$$

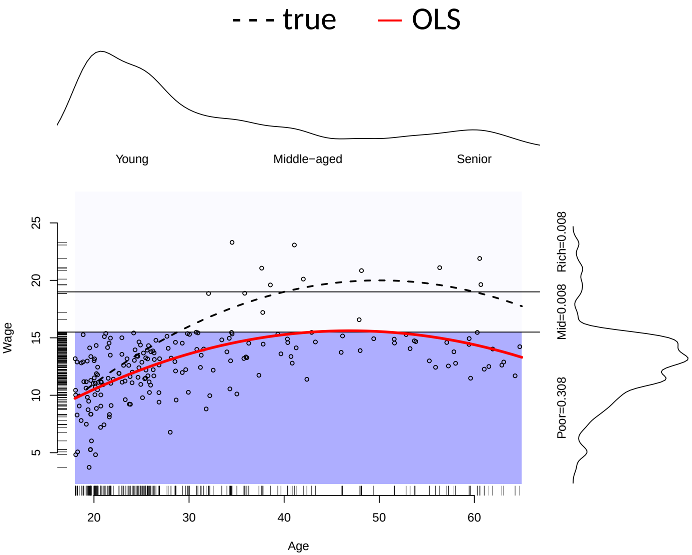
Graphic Example of Stratified Samples

Original population: $\mathbb{E}(\text{wage} \mid \text{age}) = -5 + \text{age} - 0.01\text{age}^2$

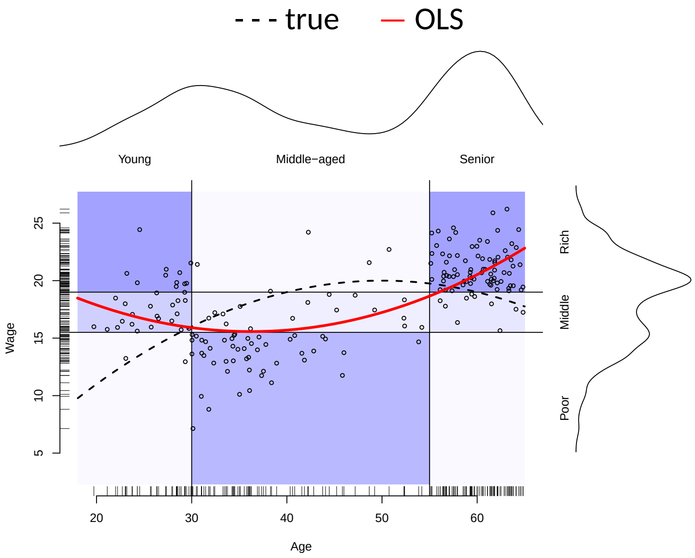



- - - true
OLS
Young=0.033
Middle−aged=0.148
Senior=0.066
Wage
5
10
15
20
25
Age
20
30
40
50
60
Rich
Middle
Poor

Stratification by Wage



Stratification by both Age and Wage



Conditional Moment Restrictions Model

- Assume a *conditional* moment restriction holds in the *target* population:
$$\mathbb{E}^*[g(Z^*, \theta^*) \mid X^*] = 0$$
- Objective: find an efficient estimator of θ^* when data are collected by VP sampling.
- Stratified sampling induces **selection bias** when the distribution is mapped from P^* to P.

- In VP sampling, the target distribution can be easily recovered from the realised distribution because

$$\mathrm{d}P(z) = \frac{b(z)}{b^*}\,\mathrm{d}P^*(z),$$

where $b(Z) \stackrel{\text{def}}{=} \sum_l p_l \mathbb{I}_{\mathbb{C}_l}(Z)$.

- Consequently,

$$\mathbb{E}^*[g(Z^*,\theta^*) \mid X^*] = 0 \iff \mathbb{E}\left[\frac{g(Z,\theta^*)}{b(Z)} \;\middle|\; X\right] = 0.$$

- Uniqueness of θ^* is not lost because $b(\cdot)$ is a known function and does not depend on any unknown parameters.
- Therefore, any model identified under P^* is identified under P.

Smooth Empirical Likelihood (SEL)

- SEL (proposed by Kitamura, Tripathi & Ahn, 2004, *Ecta*) extends the EL, a non-parametric method for testing and estimating (Owen, 1988, *Biometrika*).
- EL estimators based on unconditional moment restrictions are equivalent to optimally weighted GMM estimators.
- Parametric restrictions can be tested using a non-parametric version of Wilks' theorem (Qin and Lawless, 1994, *Ann. Stat.*). EL ratio statistics do not need to be explicitly studentised.
- SEL extends the properties of EL to estimating model characterised by conditional moment restrictions (Kitamura & Tripathi, 2003, *Ann. Stat.*), and SEL-based estimators attain the semi-parametric efficiency bounds (Severini and Tripathi, 2013).

Implementation of Our Estimator

We have independent observations $Z_1, \ldots, Z_n$, collected under VP sampling. The objective is to use them to estimate the parameter θ^* defined by the conditional moment restrictions:

$$\mathbb{E}[\rho_1(Z, \theta^*) \mid X] = 0,$$

where $\rho_1(Z, \theta^*) \stackrel{\text{def}}{=} \frac{g(Z,\theta^*)}{b(Z)}$.
In order to take into account conditioning, construct kernel weights

$$w_{ij} \stackrel{\text{def}}{=} \frac{K(X_i - X_j)}{\sum_{k=1}^{n} K(X_i - X_k)}, \quad i, j = 1, \ldots, n.$$

The SEL estimator solves the optimisation problem:

$$\max_{p_{ij}} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \log p_{ij} \quad \text{s. t.} \quad p_{ij} \geq 0, \quad \sum_{i=1}^{n} \sum_{j=1}^{n} p_{ij} = 1,$$

$$\sum_{j=1}^{n} \rho_1(Z_j, \theta) p_{1j} = 0, \ \ldots, \ \sum_{j=1}^{n} \rho_1(Z_j, \theta) p_{nj} = 0.$$

The empirical probabilities p_{ij} of each observation Z_j have the expression

$$\hat{p}_{ij}(\theta) \stackrel{\text{def}}{=} \frac{1}{n}\left(\frac{w_{ij}}{1+\lambda_i'\rho_1(Z_j,\theta)}\right), \quad i,j=1,\dots,n,$$

where $\lambda_i,\dots,\lambda_n$ are the Lagrange multipliers of the constraints. Plugging the expression for $\hat{p}_{ij}$ yields the SEL estimator of θ^*:

$$\hat{\theta}_{\text{SEL}} = \underset{\theta}{\operatorname{argmax}}\left[-\max_{\lambda_1,\dots,\lambda_n}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\log\bigl(1+\lambda_i'\rho_1(Z_j,\theta)\bigr)\right]$$

Our contribution: We extend Wooldridge (1999, *Ecta*) result on efficiency bounds in unconditional moment restrictions models under VP sampling to conditional moment restrictions models, and show that $\hat{\theta}_{\text{SEL}}$ is asymptotically efficient in the sense of Chamberlain (1987, *JoE*).

Consider the linear regression model

$$Y^* = \alpha^* + X^{*\prime}\beta^* + U^*$$

where all regressors are exogenous, i. e. $\mathbb{E}(U^* \mid X^*) = 0$.
Note that

$$\begin{aligned}\mathbb{E}(U^* \mid X^*) = 0 &\iff \mathbb{E}(Y^* - \alpha^* - X^{*\prime}\beta^* \mid X^*) = 0\\ &\iff \mathbb{E}[g(Z^*, \theta^*) \mid X^*] = 0,\end{aligned}$$

where $Z^* \stackrel{\text{def}}{=} \begin{pmatrix} Y^* \\ X^* \end{pmatrix}$ and $\theta^* \stackrel{\text{def}}{=} \begin{pmatrix} \alpha^* \\ \beta^* \end{pmatrix}$.

Estimators Compared in the Simulations

$$\hat{\theta}_{\mathrm{LS}} \stackrel{\mathrm{def}}{=} \left(\sum_{i=1}^{n} \tilde{X}_i \tilde{X}_i'\right)^{-1} \left(\sum_{i=1}^{n} \tilde{X}_i Y_i\right) \qquad (\tilde{X} \stackrel{\mathrm{def}}{=} \begin{pmatrix} 1 \\ X \end{pmatrix})$$

$$\hat{\theta}_{\mathrm{GMM}} \stackrel{\mathrm{def}}{=} \left(\sum_{i=1}^{n} \frac{\tilde{X}_i \tilde{X}_i'}{b(X_i, Y_i)}\right)^{-1} \left(\sum_{i=1}^{n} \frac{\tilde{X}_i Y_i}{b(X_i, Y_i)}\right)$$

$$\hat{\theta}_{\mathrm{SEL}} = \operatorname*{argmax}_{\theta} \left[- \max_{\lambda_1, \ldots, \lambda_n} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \log\big(1 + \lambda_i' \rho_1(Z_j, \theta)\big) \right]$$

Asymptotic Variance of Estimators

	Endogenous stratification	Exogenous stratification
LS	Not consistent	$(\mathbb{E}\tilde{X}\tilde{X}')^{-1}(\mathbb{E}\tilde{X}\tilde{X}'V_{1,\text{ex}}(X))(\mathbb{E}\tilde{X}\tilde{X}')^{-1}$
GMM	$\left(\mathbb{E}\frac{\tilde{X}\tilde{X}'}{b(Y)}\right)^{-1}(\mathbb{E}\tilde{X}\tilde{X}'V_{1,\text{end}}(X))\left(\mathbb{E}\frac{\tilde{X}\tilde{X}'}{b(Y)}\right)^{-1}$	$\left(\mathbb{E}\frac{\tilde{X}\tilde{X}'}{b(X)}\right)^{-1}(\mathbb{E}\tilde{X}\tilde{X}'V_{1,\text{ex}}(X))\left(\mathbb{E}\frac{\tilde{X}\tilde{X}'}{b(X)}\right)^{-1}$
SEL	$\left(\mathbb{E}\frac{\tilde{X}\tilde{X}'}{\gamma^{*2}(X)V_{1,\text{end}}(X)}\right)^{-1}$ (efficient!)	$\left(\mathbb{E}\frac{\tilde{X}\tilde{X}'}{V_{1,\text{ex}}(X)}\right)^{-1}$ (efficient!)

- Exogenous $\Rightarrow \text{Var}\,\hat{\theta}_{\text{SEL}} \leq \{\text{Var}\,\hat{\theta}_{\text{LS}}, \text{Var}\,\hat{\theta}_{\text{GMM}}\}$, but no ranking can be made for $\text{Var}\,\hat{\theta}_{\text{LS}}$ vs $\text{Var}\,\hat{\theta}_{\text{GMM}}$.
- Endogenous $\Rightarrow \text{Var}\,\hat{\theta}_{\text{SEL}} \leq \text{Var}\,\hat{\theta}_{\text{GMM}}$.

We consider the following design (Cragg, 1983, *Ecta*):

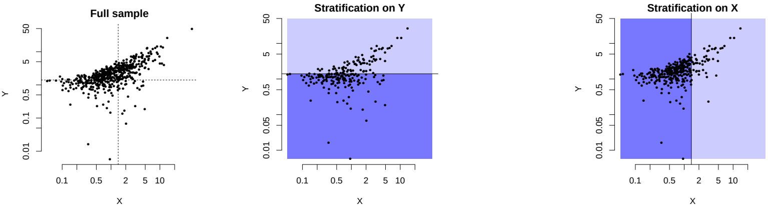
$$Y^* = \beta_0^* + \beta_1^* X^* + \sigma^*(X^*)U^*,$$

where

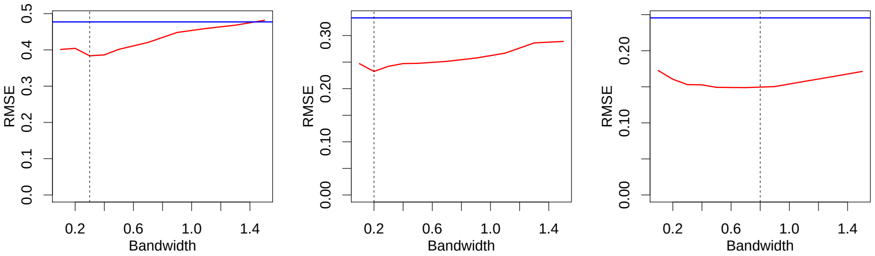
- $\theta^* \stackrel{\text{def}}{=} (\beta_0^*, \beta_1^*) = (1, 1)$
- $(U^*, \log X^*) \stackrel{\text{d}}{=} \text{NIID}(0, 1) \Rightarrow \mathbb{E}[U^* \mid X^*] = 0$
- $\sigma^*(X^*) \stackrel{\text{def}}{=} \sqrt{0.1 + 0.2X^* + 0.3{X^*}^2}$

Comparing the Designs

Stratification	Endogenous	Exogenous
$\mathbb{C}_1$	$(-\infty,\infty)\times(-\infty,1.4)$	$(-\infty,1.4)\times(-\infty,\infty)$
$\mathbb{C}_2$	$(-\infty,\infty)\times[1.4,\infty)$	$[1.4,\infty)\times(-\infty,\infty)$
(p_1,p_2)	$(0.9,0.3)$	$(0.9,0.3)$

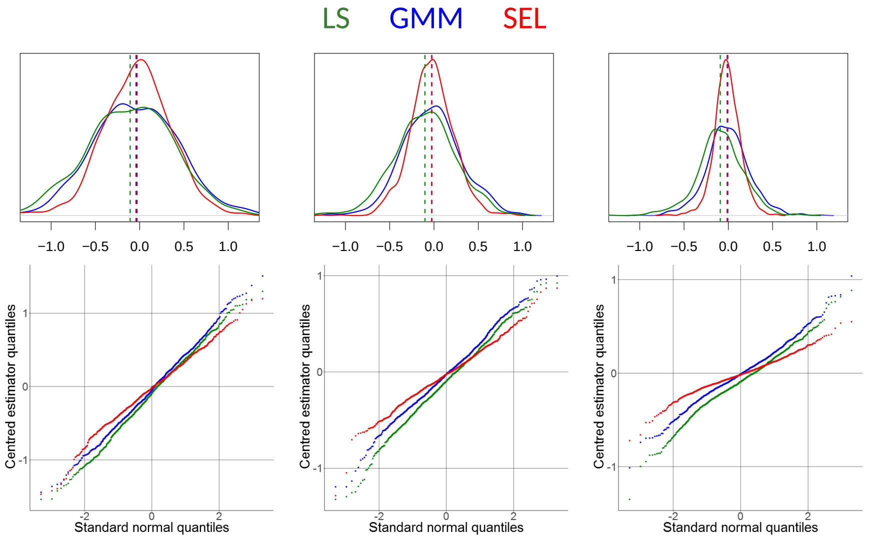


GMM SEL



$n = 50$ (≈ 24 real) $n = 150$ (≈ 71 real) $n = 500$ (≈ 235 real)

Densities and Quantiles of Centred Estimators



$n = 50$ (≈ 24 real) $n = 150$ (≈ 71 real) $n = 500$ (≈ 235 real)

- All simulations were performed on the HPC cluster of the University of Luxembourg.
- The **R** code is freely and openly available on GitHub at `https://github.com/Fifis/SELshares`.
- The non-linear nature of SEL estimator and the non-existence of a closed-form expression can present numerical challenges.
- Our implementation can estimate models on data collected under VP sampling (with or without estimation of aggregate shares).

Future Research

- As with many semi-parametric methods, there are bandwidth issues.
- There are efficiency gains if the kernel weights w_{ij} incorporate information about the distribution of X.
  - $\Rightarrow$ There must be a data-driven way to pick the optimal SEL bandwidth.
  - $\Rightarrow$ There must be a transformation of X's that leads to efficiency gains.
- In progress: Extending SEL to models with conditional moment restrictions where some observations are missing.

- Introduced a class of estimators based on SEL for models defined by conditional moment restrictions under VP sampling.
- Compared theoretically the efficiency properties of SEL, GMM and LS estimators of the parameters of a linear regression model and the aggregate shares under VP.
- Carried out a Monte Carlo experiment to check the theoretical predictions.
- For the parameters of the linear regression model, SEL has lower variance than the competitors under heteroskedasticity.

The End

Thank you for your attention!

Any questions?