Estimating counterfactual distributions through reweighting methods

Laurent Donzé
University of Fribourg (Switzerland), Department of Quantitative Economics
Bd de Pérolles 90
CH – 1700 Fribourg, Switzerland
E-mail: Laurent.Donze@UniFr.ch

ABSTRACT

Counterfactual distributions are generally estimated by matching techniques, less frequently by data reweighting. In this study, we investigate this second approach, first by a Monte Carlo study and then by an empirical application to Swiss wages. The main question is to determine if the estimated weights — by a probit model or nonparametrically — are really able to replicate the counterfactual distribution. As the true counterfactual distribution will be never known, a definitive answer is not possible. Nevertheless, thank to the Monte Carlo study a sensitivity analysis shows how a model or a nonparametric approach performs the task. Finally, several counterfactual Swiss wages distributions are estimated and compared each other in order to put in evidence some misspecification errors or biases.

Keywords: matching methods, counterfactual wages distribution, test of distribution, Series Logit Estimator

1. Counterfactual distribution estimation by reweighting

In many statistical problems, e.g. program or treatment evaluations, one needs to evaluate counterfactual distributions. Matching techniques or reweighting procedures can be applied. In this study, we investigate the second one approach to estimate counterfactual wage distributions. In fact, the reweighting approach (inverser probability weighting) is suggested e.g. by Firpo et al. (2007). It appears to be a simple and efficient method, which can be easily applied in the case of wages' analysis. The weights to be found are based on the concept of propensity scores. So our problem is essentially to find an unbiased and efficient estimation of these scores.

We consider two mutually exclusive groups of workers A and B. For each unit i, one has $D_{Ai} + D_{Bi} = 1$, where $D_{gi} = \mathbb{I}\{i \text{ is in } g\}$, $g = A, B$, and $\mathbb{I}\{\cdot\}$ is an indicator function. Let Y_{gi}, $g = A, B$, the wage of unit i in group g. We suppose that the wage is related to a set of covariates $\boldsymbol{X}$ by the following general functional form: $Y_{gi} = m_g(\boldsymbol{X}_i, \epsilon_i)$, where ϵ is a noise component. Let $m^C(\cdot,\cdot)$ the counterfactual wage structure. We assume that $m^C(\cdot,\cdot) \equiv m_A(\cdot,\cdot)$ for the units of group B and $m^C(\cdot,\cdot) \equiv m_B(\cdot,\cdot)$ for the units of group A. The distribution $F_{Y_g|D_g}$ may be written as:

$$F_{Y_g|D_g}(y) = \int F_{Y_g|\boldsymbol{X},D_g}(y \mid \boldsymbol{X} = \boldsymbol{x}) \cdot \mathrm{d}F_{\boldsymbol{X}|D_g}(\boldsymbol{x}), \quad g = A, B. \tag{1}$$

One can find the counterfactual distribution $F_{Y_A^C:\boldsymbol{X}=\boldsymbol{x}|D_B}$ as:

$$F_{Y_A^C:\boldsymbol{X}=\boldsymbol{x}|D_B} = \int F_{Y_A|\boldsymbol{X},D_A}(y \mid \boldsymbol{X} = \boldsymbol{x}) \cdot \mathrm{d}F_{\boldsymbol{X}|D_B}(\boldsymbol{x}). \tag{2}$$

One can rewrite the expression (2) as:

$$F_{Y_A}^C(y) = \int F_{Y_A|X_A}(y \mid \boldsymbol{X})\Psi(\boldsymbol{X})\mathrm{d}F_{X_A}(\boldsymbol{X}), \tag{3}$$

where

$$\Psi(\boldsymbol{X}) = \mathrm{d}F_{X_B}(\boldsymbol{X})/\mathrm{d}F_{X_A}(\boldsymbol{X}) \tag{4}$$

is the reweighting factor. It is obvious in (3) that the counterfactual is obtained by reweighting the distribution F_{Y_A}. The weighting factor $\Psi(\boldsymbol{X})$ defined in (4) is a ratio of two multivariate marginal distributions of the covariates $\boldsymbol{X}$. Indeed, one has

$$\Psi(\boldsymbol{X}) = \frac{\Pr(\boldsymbol{X} \mid D_B = 1)}{\Pr(\boldsymbol{X} \mid D_B = 0)} = \frac{\Pr(D_B = 1 \mid \boldsymbol{X})/\Pr(D_B = 1)}{\Pr(D_B = 0 \mid \boldsymbol{X})/\Pr(D_B = 0)}. \tag{5}$$

The expression (5) suggests the following procedure to estimate Ψ. One has to estimate by a probit or logit model the probabilities $\Pr(D_B = 1 \mid \boldsymbol{X})$ and $\Pr(D_B = 0 \mid \boldsymbol{X})$. The probabilities $\Pr(D_B = 1)$ and $\Pr(D_B = 0)$ are simply estimated by the sample proportions of each group.

2. Efficient propensity scores estimation

The probability to participate to a program, or to be treated, conditionally to a set of covariates is called a propensity score. Thus the probabilities $\Pr(D_B = 1 \mid \boldsymbol{X})$ and $\Pr(D_B = 0 \mid \boldsymbol{X})$ can be assimilated to a propensity score. One has just above mentioned that we can estimate these probabilies by a probit or logit model. Hirano et al. (2003) propose a sieve approach by the so–called Series Logit Estimator (see e.g. Geman and Hwang (1982)). Based on the idea of approximating an unknown function by a polynomial function, sequences of power series are used to model the structural function of a logit regression model.

Let $w_C(D_B, \boldsymbol{X})$ be the weighting function. One can easily estimate it by:

$$\hat{w}_c(D_B, \boldsymbol{X}) = \frac{1 - D_B}{\hat{p}} \left(\frac{\hat{p}(\boldsymbol{X})}{1 - \hat{p}(\boldsymbol{X})} \right), \tag{6}$$

where $\hat{p} = N^{-1} \sum_{i=1}^{N} D_{Bi}$, N is the total number of observations, and $\hat{p}(\boldsymbol{X})$ is the logit / probit estimation of the propensity score. One can normalize $\hat{w}_c(D_B, \boldsymbol{X})$ by $\hat{w}_c^*(D_{Bi}, \boldsymbol{X_i}) := \hat{w}_c(D_{Bi}, \boldsymbol{X_i}) / \sum_{i=1}^{N} \hat{w}_c(D_{Bi}, \boldsymbol{X_i})$.

We estimate the propensity scores with several functional form specifications, generate the corresponding weights and finally compute the counterfactual distributions. A sensitivity analysis is made in the case of the Swiss wages.

REFERENCES

[1] Sergio Firpo, Nicole Fortin, and Thomas Lemieux. Decomposing wage distributions using recentered influence function regressions. Technical report, PUC-Rio, UBC, and National Bureau of Economic Research, 2007.

[2] Stuart Geman and Chii-Ruey Hwang. Nonparametric maximum likelihood estimation by the method of sieves. *The Annals of Statistics*, 10(2):pp. 401–414, 1982. ISSN 00905364. URL `http://www.jstor.org/stable/2240675`.

[3] Keisuke Hirano, Guido W. Imbens, and Geert Ridder. Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica*, 71(4):pp. 1161–1189, 2003. ISSN 00129682. URL `http://www.jstor.org/stable/1555493`.