World Academy of Science, Engineering and Technology
International Journal of Economics and Management Engineering
Vol:8, No:8, 2014

On Estimating the Headcount Index by Using the Logistic Regression Estimator

Encarnación Álvarez, Rosa M. García-Fernández, Juan F. Muñoz, Francisco J. Blanco-Encomienda

Abstract—The problem of estimating a proportion has important applications in the field of economics, and in general, in many areas such as social sciences. A common application in economics is the estimation of the headcount index. In this paper, we define the general headcount index as a proportion. Furthermore, we introduce a new quantitative method for estimating the headcount index. In particular, we suggest to use the logistic regression estimator for the problem of estimating the headcount index. Assuming a real data set, results derived from Monte Carlo simulation studies indicate that the logistic regression estimator can be more accurate than the traditional estimator of the headcount index.

Keywords—Poverty line, poor, risk of poverty, sample, Monte Carlo simulations.

I. Introduction

THE problem of estimating a proportion has important applications in the field of economics, and in general, in many areas such as social sciences. Indeed, many social indicators are based upon binary variables or require the use of proportions to obtain such indicators. The application discussed in this paper is the estimation of the headcount index, i.e., the proportion of individuals falling below the official poverty line. In general, the relative poverty line is obtained by using indicators based on variables such as income or expenditures. Broadly speaking, it is set a threshold below which people are classified as poor. To obtain this level normally is fixed a percentage of the median which can be 50%, 60% or even 30% when severe poverty is considered.

Note that this poverty indicator (the headcount index) is widely used in comparison of poverty across countries. For instance, Eurostat fixes the poverty line in the 60% of the median of the equivalised net income. The latest results published by Eurostat based on the EU-Statistics on Income and Living Conditions 2012 show that the proportion or people at risk of poverty is 17% in the Euro area (17 countries) compared with 22.2% in Spain and 16.1% in Germany. The OECD defines income poverty as the share of people living in equivalised household income with less than 50% of the median equivalised household disposable income. The poverty rate for OECD-20 was equal to 11.1 % in 2010 compared with 15.4% in Spain and 8.8% in Germany. Relevant references related to poverty indicators are [3], [4], [6].

The fight again poverty is in a priority place in the political agendas. For instance, to eradicate the extreme poverty is the first of the Millennium Development Goals. We can find another example in Europe 2020 strategy according to which at least 20 million people should lift out on poverty and social exclusion. This important target requires, as discussed as [7], the use of more accurate information related to the living conditions of people can be a useful tool to study and reduce poverty. In this sense, the main aim of this paper is to define new quantitative methods for estimating the headcount index, also named as the proportion of people at risk of poverty.

The logistic regression estimator was introduced by [5]. This estimator assumes a logistic model and depends on the parameter associated to the logistic model. Algorithms, such as Newton-Raphson, are generally used to estimate the model by maximizing likelihood function numerically. References related to the estimation of a proportion can be seen by [1], [2], [9], [13], [14].

In Section II we define the headcount index as a proportion, and the aim is to estimate this proportion by using the logistic regression estimator, which is defined in Section III. A real data set is described in Section IV. These data are considered as a population, from which samples are obtained in order to carry out a Monte Carlo simulation. Results derived from this study indicate that the logistic regression estimator can be more accurate than the traditional estimator of the headcount index.

II. The Headcount Index

Let $U = \{1, \ldots, N\}$ be a finite population of N individuals. We consider the problem of estimating the headcount index, which can be defined as

$$P = \frac{N_p}{N}$$

where N_p is the number of poor in the population or the number of people at risk of poverty.. Let y be a suitable quantitative variable of welfare, such as income or expenditure. A poverty line (L) is commonly used by the statistical agencies to classify the population into poor and non-poor, i.e., an individual is considered as poor if its income or expenditure is less than the poverty line. Assuming this scenario, we can calculate the number of poor as $N_p = \sum_{i \in U} A_i$, where $A_i = \delta(y_i \leq L) = 1$ if the ith individual is classified as poor and $A_i = 0$ otherwise. $\delta(\cdot)$ is the indicator variable, which

E. Álvarez is with the Department of Quantitative Methods in Economics and Business, University of Granada, Granada, CP 18071, Spain (e-mail: encarniav@ugr.es).

R.M. García-Fernández is with the Department of Quantitative Methods in Economics and Business, University of Granada, Granada, CP 18071, Spain (e-mail: rosamgf@ugr.es).

J.F. Muñoz is with the Department of Quantitative Methods in Economics and Business, University of Granada, Granada, CP 18071, Spain (e-mail: jfmunoz@ugr.es).

F.J. Blanco-Encomienda is with the Department of Quantitative Methods in Economics and Business, University of Granada, Granada, CP 18071, Spain (e-mail: jble@ugr.es).

takes the value 1 if its argument is true and 0 otherwise. Note that the population proportion P can be seen as the mean of the attribute of interest A with values given by $A_1, \ldots, A_N$. In this paper, we assume that the poverty line (L) is established by the corresponding authority, i.e., L is fixed at some official quantity. Note that statistical agencies usually define the poverty line as the 60% of the median income. Note that the definition and calculation of the poverty line is beyond the scope of this paper.

The parameter P can be estimated by using the information collected from a sample s, with size n, and selected from the population U. We assume the sample is selected according to a probability sampling design with sampling weights given by $d_i = 1/\pi_i$, where π_i is the first-order inclusion probability for the ith individual in the population.

The traditional estimator of P is the Horvitz-Thompson estimator, which is given by

$$\widehat{p} = \frac{1}{N}\sum_{i\in s} d_i A_i.$$

In the case the sample is selected under the classical simple random sampling without replacement, the Horvitz-Thompson estimator is simply given by $\widehat{p} = n_p/n$, where $n_p = \sum_{i\in s} A_i$ is the number of poor in the sample s.

In addition to the variable of interest y, many official surveys about living conditions contain additional variables, which could be used at the estimation stage to improve the estimation of parameters. Note that some of these auxiliary variables could be observed at the population level. Let x be a quantitative auxiliary variable related to y. In this situation, the logistic regression estimator (see [5]) could be used to estimate the headcount index. This application is discussed in Section III.

III. Estimating the Headcount Index with the Logistic Regression Estimator

We now introduce the logistic regression estimator and apply this estimator to the problem of estimating the headcount index.

The logistic regression estimator is given by

$$\widehat{p}_{Lgreg} = \frac{1}{N}\left(\sum_{i\in U}\widehat{\mu}_i + \sum_{i\in s} d_i(A_i - \widehat{\mu}_i),\right)$$

where

$$\widehat{\mu}_i = \frac{\exp(x_i\widehat{\theta})}{1+\exp(x_i\widehat{\theta})}, \quad i = 1, \ldots, N,$$

are the predicted values based upon the logistic regression model

$$Pr(A_i = 1) = \quad \mu_i = \frac{\exp(x_i\theta)}{1+\exp(x_i\theta)}$$

$$Pr(A_i = 0) = \quad 1-\mu_i = 1 - Pr(A_i = 1)$$

The maximum likelihood method can be used to estimate the parameter θ. The corresponding likelihood function is given by (see, also, [5], [12])

$$L(\theta) = \Pi_{i\in U_1}\mu_i \Pi_{i\in U_0}(1-\mu_i),$$

where $U_1 = \{i : i \in U \text{and } A_i = 1\}$ and $U_0 = U - U_1$. The Newton-Raphson algorithm can be used to find the solution to the maximization problem.

Note that alternative estimators of a proportion are defined by [8], [10] and [11]. However, such estimator cannot be applied in this situation, since they assume that the auxiliary variable are also given binary variables.

IV. Monte Carlo Simulations

Assuming a Monte Carlo simulation study, we now compare the empirical performance of the logistic regression estimator to the traditional estimator of the headcount index. For this purpose, we considered a real data set selected from the 2011 Spanish Survey of Living Conditions. These data were considered as a population, and then $D = 1000$ samples were selected under simple random sampling without replacement. The aim in this study was to estimate the population headcount index, which is defined as the 60% of the median income.

Estimators of the headcount index and described in this paper are compared in terms of Relative Bias (RB) y and Relative Root Mean Square Error (RRMSE), where

$$RB = \frac{E[\widehat{p}] - P}{P} \quad ; \quad RRMSE = \frac{\sqrt{MSE[\widehat{p}]}}{P},$$

and where $\widehat{p}$ is a given estimator of P and $E[\cdot]$ and $MSE[\cdot]$ are the empirical expectation and the empirical mean square error based upon the $D = 1000$ samples, and which are given

$$E[\widehat{p}] = \frac{1}{D}\sum_{d=1}^{D}\widehat{p}(d) \quad ; \quad MSE[\widehat{p}] = \frac{1}{D}\sum_{d=1}^{D}(\widehat{p}(d) - P)^2,$$

TABLE I
Values of $RRMSE$ for the various estimators of the headcount index P. The sampling fraction is given by $f = n/N$.

	Sampling fraction f (%)			
Estimator	1	5	10	20
$\widehat{p}$	19.3	16.7	12.6	8.2
$\widehat{p}_{Lgreg}$	17.9	15.3	11.4	7.4

Values of RB derived from this study are all less than 1%. This implies that estimators have reasonable biases, and for this reason values of RB are omitted. Values of $RRMSE$ can be seen in Table I. We observe that the logistic regression estimator $\widehat{p}_{Lgreg}$ is more accurate than the traditional estimator $\widehat{p}$. This property is satisfied for the various values of the sampling fraction.

V. Conclusion

In this paper, the headcount index was expressed as a proportion. In the literature, many quantitative methods can be used for the problem of estimating a given parameter. However, the logistic regression estimator is a very common and popular technique used when the variable of interest is a binary variable. This is the situation of the headcount index.

We introduce the logistic regression estimator and show that this method can be applied to the problem of estimating the headcount index.

A simulation study based upon a real data set was carried out to analyze the precision of estimators of the headcount index. Results derived from this study indicate that the logistic regression estimator is more accurate than the traditional estimator of the headcount index.

ACKNOWLEDGMENT

This work is supported by the project (grant) P11-SEJ-7090 of the Consejería de Innovación, Ciencia y Empresa (Junta de Andalucía).

REFERENCES

[1] Y.G. Berger and C.J. Skinner, "Variance estimation for a low income proportion". *Journal of the Royal Statistical Society, Series B*, 52, pp. 457–468 2003.
[2] C.R. Blyth and H.A. Still, "Binomial confidence intervals". *Journal of the American Statistical Association*, 78, pp. 108–116, 1983.
[3] E. Crettaz and C. Suter, "The impact of Adaptive Preferences on Subjective Indicators: An Analysis of Poverty Indicators". *Social Indicators Research*, 114, pp. 139-152, 2013.
[4] F. Giambona, F. and E. Vassallo, "Composite Indicator of Social Inclusion for European Countries". *Social Indicators Research*, 116, pp. 269-293, 2014.
[5] R.Lehtonen and A. Veijanen, "On multinomial logistic generalized regression estimators", *Survey Methodology*, 24, pp. 51-55, 1998.
[6] M. Medeiros, "The Rich and the Poor: the Construction of an Affluence Line from the Poverty line". *Social Indicators Research*,78, pp. 1-18, 2006.
[7] I. Molina and J.N.K. Rao, "Small area estimation of poverty indicators", *The Canadian Journal of Statistics*, 38, pp. 369-385, 2010.
[8] J.F. Muñoz, E. Álvarez, A. Arcos, M.M. Rueda, S. González and A. Santiago, "Optimum ratio estimators for the population proportion", *International Journal of Computer Mathematics*, 89, pp. 357-365, 2012.
[9] R.G. Newcombe, "Two-sided confidence intervals for the single proportion: comparison of seven methods". *Statistic in Medicine*, 17, pp. 857–872, 1998.
[10] M.M. Rueda, J.F. Muñoz, A. Arcos, E. Álvarez and S. Martínez, "Estimators and confidence intervals for the proportion using binary auxiliary information with application to pharmaceutical studies". *Journal of Biopharmaceutical Statistics*, 21, pp. 526–554, 2011.
[11] M.M. Rueda, J.F. Muñoz, A. Arcos and E. Álvarez, "Indirect estimation of proportions in natural resource surveys". *Mathematics and Computers in Simulation*, 81, pp. 2317–2325, 2011.
[12] C.E. Särndal, B. Swensson and J. Wretman, *Model Assisted Survey sampling*, Springer Verlag, 1992.
[13] S.E. Vollset, "Confidence interval for a binomial proportion". *Statistic in Medicine*, 12, pp. 809–824, 1993.
[14] E.B. Wilson, "Probable inference, the law of succession, and statistical inference". *Journal of the American Statisitical Association*, 22, pp. 209–212, 1927.

Encarnación Álvarez is a lecturer in the Department of Quantitative Method in Economics and Business at the University of Granada in Granada, Spain. Her research is about the estimation of proportions and applications in poverty.

Rosa M. García-Fernández is associate professor in the Department of Quantitative Method in Economics and Business at the University of Granada in Granada, Spain. Her research is about the analysis and study of the poverty and inequality.

Juan F. Muñoz is associate professor in the Department of Quantitative Method in Economics and Business at the University of Granada in Granada, Spain. His research is about quantitative methods used for the estimation of parameter.

Francisco J. Blanco-Encomienda is associate professor in the Department of Quantitative Method in Economics and Business at the University of Granada in Granada, Spain. His research is about quantitative methods in economics and business.