IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728, p-ISSN: 2319-765X. Volume 14, Issue 1 Ver. I (Jan. - Feb. 2018), PP 62-64
www.iosrjournals.org

Application of the Robust Regression Model in Election of African Countries

Edwin Ayora[1], RomanusOdhiambo[2], George Orwa[3], Joseph Mung'atu[4]
Researcher,Jomo Kenyatta University of Agriculture and Technology
Professor of statistics,Jomo Kenyatta University of Agriculture and Technology
Professor of statistics,Jomo Kenyatta University of Agriculture and Technology
Doctor of statistics,Jomo Kenyatta University of Agriculture and Technology
Corresponding Author: Edwin Ayora

Abstract: *Robust regression model has proved to provide efficiency in presence of errors with heavy tailed distribution. The application of the robust regression has gained a lot of application because of its indiscriminate in either parametric or non-parametric methods. This paper examines application of modified weighting method to a cross national regression analysis of election data in Africa countries. Data for analysis was obtained from African Social Survey(ASS)2002/2003 , the data covered 48 countries though only 29 countries were used because of comparability of survey questions. The results obtained reveals* X_{wx}^2 *to be higher for the weights within-country. The paper does not observe clearly the effects expected from misspecification might simply mean that the model is adequately specified. The 95 critical value for a chi-squared with 5 d.f. is 11.1 and, whilst recognizing that this is a very crude approximation to its null distribution* X_{wx}^2 *only 1 of the 27 countries .Statistically speaking the weights for the model obtained from the design weights and within the country weights have no significant difference hence essential in the model. Thus, robust regression models is a sure way for modelling the election turnout in countries*

Key words: *Robut regression model, Election survey. design weights, within the country weights*

Date of Submission: 09-01-2018 Date of acceptance: 25-01-2018

I. Introduction

The expression"robustregression"de notesaset of estimation techniques that are less sensitive than ordinary leas tsquares(OLS) to the effect of possible influential observations. The main argument invoked to justify the use of robust regression is that it provides efficiency gains in the presence of errors with heavy taileddistributions.Robustregressionhasawell-establishedtraditioninstatistics. Nevertheless, over the past decade, a form of robust regression based on Huber's M-estimator was made available in popular soft ware packages. The particular version of this estimator that has become popular in applied econometrics is based on thealgorithmproposedbyLi(1985),whichisaniterativelyreweightedleastsquares algorithm usingbi weights(Beaton and Tukey,1974). However, perhapsbe cause of the lack of appropriatereferencesonitsusemostpractitioners who have used this estimator seem to be unaware of the fact that its properties depend on strong assumptions about the symmetry an dhomoscedasticity of the errors, an djustifyitsuse withmisleadingclaimsaboutitsadvantages.Inparticularinthepresenceofskewed heteroscedastic errors this M-estimator will be inconsistent for these parameters and note that its efficiency can be severely affected by heteroskedasticity. The paper explores the application of modified weighting method to a specific cross national regression analysis of Kenya and Zimbabwe survey data. The application exemplifiesaparticularproblem of weight in garisingincrossnational comparative surveyswhendataarepooledacrosscountries(Thompson,2008).Itiscommonin the design of such surveys for sample sizes in different countries to be much less variable than population sizes and this leads to very different sampling fractionsacrosscountries.Thisimpliesthatdatafromlargecountriesmaydominatepooled analyses employing Horitz Thompson weighting, leading to inefficient use of sample data (Thopmson, 2008).

II. Methodology

Fieldhouse et al. (2007) describe as a rational choice model and which they accounts well for country-level variation in comparison with two other models. The model includes two sets of explanatory variables: variable of political science interest, which reflect voting behaviour as a rational choice, and basic demographic control variables. The variables of political science interest include five scales derived from principal component analysis of individualquestions: the first two principal components of questions measuring the extent to which respondents think they can understand and influence politics, termed 'political efficacy 1 and2'; the first two principal components of questions measuring respondents' feelings of civic duty, termed 'system benefits 1 and 2'; and the first principal component of questions relating to satisfaction with the economy, government competence, democracy, education and health services, termed collective benefits'. In addition there is a measure of partisanship and of the closeness of the contest (the difference in vote between the first and second placed parties in the election) and the interaction between these two variables. The five demographic control variables consist of gender; whether the respondent belongs to an ethnic minority; whether the respondent was born in the country; whether the respondent has a partner; and whether the respondent has a child.

The data comes from ASS in 2002/2003. Although the survey covered 48 countries, the paper uses data from 29 countries because of concern about the comparability of survey questions pertinent to this analysis. Of the 42,359 respondents of the ASS, 3,549 from the 19 countries were removed leaving 38,310 respondents. Of, these, 3,787 were aged 18-24 years old and, of these, 3,109 were eligible to vote. Among these 3,109 respondents 488 had missing data for at least one of the variables used in the model. This left 2,621 complete records which were used in the analysis. There was no obvious systematic reason for the item nonresponse. There are many variables underlying this analysis, since several variables in the model are principal components of other variables and all of the attitudinal variables had some amount of missing data. There was much less missing data in the socio-demographic control variables and, perhaps most importantly, there were only 5 missing values for the outcome variable on whether the respondent voted. The variable with the most severe item nonresponse was satisfaction with government with 130 missing values. This paper did not attempt to take accountof any potential biasing effects from this item nonresponse nor from unit nonresponse in the ASS.

Samples in the ASS were selected independently in different countries, each with a minimum effective sample size 1500 (or 800 if the country had fewer than 10 million population) In Some countries population registers were used to select individuals by single stage sampling with equal probability while in other countries different forms of stratified multi-stage sampling were employed with varying kinds of strata, multi-stage units and numbers of stages (African Social Survey, 2004). In these countries, the probabilities of inclusion of individuals could vary for a number of reasons. Just one individual was typically selected per sampled household or address so that inclusion probabilities would vary by the number of eligible individuals in the household or address. Other reasons for the inclusion probabilities to vary included: differential sampling fractions between strata; the use of probability proportional to size sampling in some countries to select multistage units; and different sampling procedures for individuals with and without listed telephone numbers.

III. Results and Discussion

To explore the effect of the choice of weights on model fit, the observations were divided in each country into weighted quintile groups$k = 1,\ldots..5$, according to the values of the probability $\hat{p}_{wj}$ that $Y_j = 1$, predicted using the model estimated with weighting method w, following the approach of Graubard et al. (1997). For each country$x$, weight $w$and quantile group Z computed $p_{wxz}|$, the weighted observed proportion with $y_j = 1$, and $\hat{p}_{wxz}$ the weighted mean of the $\hat{p}_{wj}$. As a simple measure of fit in country x of the model estimated using weights w, computed.

$$A_{wx} = \frac{\sum_{k=1}^{5} |p_{wxz} - \hat{p}_{wxz}|}{5}$$

Since sample size varies quite considerably between countries (Table 5) and smaller sample sizes may tend to inflate A_{wx} irrespective of the validity of the model, Also,

$$X_{wx}^2 = \frac{\sum_{k=1}^{5} (p_{wxz} - \hat{p}_{wxz})^2}{p_{wxz}\,(1-\hat{p}_{wxz})/n_{wxz}}$$

Based on the test statistic of Hosmer and Lemeshow (1980), where n_{wxz} is the sample size in the quintile group. The statistics A_{wx} and X_{wx}^2 are designed primarily for comparative purposes, but, as a very crude test of fit and compare X_{wx}^2 against critical values of a chi-squared distribution with 5 degrees of freedom. This is very crude because it takes no account of the sampling variation in the fitted values$\hat{p}_{wj}$ nor of the weighting. Alternative approaches which take account of the weighting have been proposed (Graubard et al., 1997; Archer et al., 2007) but these would require extension to assess fit in subpopulations, such as countries in this paper

Values of A_{wx} and X_{wx}^2 are presented in *Table 4.1* by country, for the design weights and the within-country weights. The rows are ordered by the value ofA_{wx} for the design weights. About model misspecification, the research anticipates that the fit for countries with high between-country weights, such as South Africa and Nigeria, might be better with the design weights. This is true for South Africa when countries are ranked by either A_{wx} and X_{wx}^2 measures, but not for Nigeria. Conversely, the country with the smallest between-country weight, Madagascar, fit better with the within-country weights than the design weights. This is true for X_{wx}^2 but not for A_{wx} with respect to which Madagascar has the worst fit with either choice of weights. Another pattern which might be anticipated from the discussion of model misspecification is that goodness of fit may show greater dispersion between countries for the design weights than for the within-country weights. There seems little evidence of this. The standard deviation of A_{wx} between countries is almost identical for the two types of weights. For X_{wx}^2 it is infact higher for the within-country weights. The fact that the paper does not observe clearly the effects expected from misspecification might simply mean that the model is adequately specified. The 95 critical value for a chi-squared with 5 d.f. is 11.1 and, whilst recognizing that this is a very crude approximation to its null distribution X_{wx}^2 only 1 of the 27 countries is above this figure, just as would be anticipated if the model were true.

Country	Sample Size	Design Weights		Within Country Weights	
		A_{wx}	X_{wx}^2	A_{wx}	X_{wx}^2
Nigeria	200	0.02	0.79	0.03	1.80
Swaziland	97	0.04	1.51	0.03	1.34
Egypt	208	0.04	2.42	0.05	3.81
Ghana	112	0.04	2.00	0.03	0.65
Uganda	138	0.05	1.63	0.09	6.52
Algeria	149	0.05	4.51	0.05	4.86
Botswana	84	0.05	3.55	0.04	2.08
Djibouti	58	0.06	2.15	0.06	2.64
Ethiopia	224	0.06	4.89	0.08	9.94
Gabon	90	0.06	2.31	0.05	1.38
Mozambique	102	0.06	2.52	0.08	4.47
Sudan	129	0.06	3.76	0.06	3.61
Morocco	116	0.07	6.07	0.04	2.57
Gambia	92	0.08	3.58	0.08	5.27
Equatorial Guinea	52	0.08	2.01	0.10	2.58
Tanzania	156	0.08	6.74	0.05	2.10
Kenya	149	0.08	9.62	0.10	14.57
Democratic Republic of Congo	169	0.09	10.20	0.09	8.12
South Africa	160	0.10	12.63	0.10	11.04
Lesotho	76	0.14	10.49	0.11	8.19
Cote d'Ivoire	101	0.14	2.51	0.05	4.92
Madagascar	99	0.14	2.31	0.02	1.35
Cameroon	112	0.14	6.12	0.03	2.56
Malawi	121	0.15	2.56	0.03	3.31
Niger	110	0.16	0.04	2.10	0.04
Burkina Faso	93	0.16	2.34	0.07	1.39
Mali	131	0.19	3.75	0.09	3.43
Zambia	98	0.19	1.51	0.02	1.91
Angola	107	0.19	9.31	0.23	6.31

References

[1]. Fieldhouse, E, Tranmer, M., and Rusell, A (2007). Electoral participation of youngpeopleinEurope.
[2]. Hosmer,D.W.andLemeshow,S.(1980)Goodness-of-fittestsforthemultiple 1. logistics regression model
[3]. Thompson, M. E. (2008). International Surveys, Motives and methodologies.
[4]. African Social Survey (2004). ASS Round one 2002/2003
[5]. Graubard,B.I.,Korn,E.L.andMidthane,D.(1997).Testinggoodness-of-fit forlogisticregressionwithSurveydata.

Edwin Ayora "Application of the Robust Regression Model in Election of African Countries." IOSR Journal of Mathematics (IOSR-JM) 14.1 (2018): 62-64.