Lithuanian Journal of Statistics *2015, vol. 54, No 1, pp. 45–51*
Lietuvos statistikos darbai *2015, 54 t., Nr. 1, 45–51 p.*
www.statisticsjournal.lt

TWO-WAY ANOVA FOR THE STUDY OF REVENUE MOBILIZATION INEQUALITIES

Kwaku F. Darkwah, Richard Tawiah[1], Maxwell Adu-Gyamfi

Dept. of Mathematics, Kwame Nkrumah University of Science and Technology
Address: Dept. of Mathematics, College of Science, KNUST, Kumasi, Ghana
E-mail: [1]rtawiah64@yahoo.com

Received: March 2015 Revised: June 2015 Published: November 2015

Abstract. The Ghanaian informal sector is characterized by underemployment, bad working conditions, uncertain work relationships and low earnings. This sector employs 80% of the workforce. Consequently, establishing an efficient tax system that can raise sufficient revenue to finance essential expenditures without recourse to government borrowing has been a challenge. Using time series data extracted from a revenue mobilization unit in Ghana, this paper examines the inequalities in four revenue mobilization instruments, namely self-employed tax, company tax, Pay As You Earn (PAYE) tax and miscellaneous taxes. The study uses the theory of two-way ANOVA as the main approach for exploring the differences. We assessed the adequacy of our theoretical approach using numerical methods. Revenues generated from the mobilization instruments differed significantly with exception of self-employed tax and company tax. Generally, revenue from company, self-employed and miscellaneous taxes were low, compared to Pay As You Earn (PAYE) tax. This study draws attention to the importance of enhancing the informal sector in order to improve revenue from self-employed tax.

Keywords: Analysis of variance, Ghana, Revenue Mobilization Instrument, Taxation.

1. Introduction

Taxation is a legal framework through which governments collect revenue from its citizenry. It provides governments with the funds required to invest in development, relieve poverty and deliver public services. In Ghana, revenue mobilization is done through direct and indirect tax administration. Direct tax is one which is intended to be paid by a person or organization on whom or which it is actually levied [1]. However, indirect tax is a tax which is levied on a single person or organization with the expectation that the tax will be shifted or passed on to another [1]. Revenue that accrues from direct tax consists of income tax, stamp duty, gift tax, capital gains tax and other special taxes such as national fiscal stabilization levy and airport tax. Income tax consists of corporate tax, personal income tax (Self-Employed tax) and Pay-As-You-Earn (PAYE). Corporate tax is a tax paid by companies on their profits in the year. Personal income tax which is also called self-employed tax requires persons or individuals to pay income tax at graduated rates in four equal installments. The Pay-As-You-Earn (PAYE) is a means of paying tax on incomes earned by employees [2].

A major concern in public economics is the design of efficient tax system. For developing countries an efficient tax has been explained by Tanzi and Zee [3] as the tax that can raise essential revenue without excessive government borrowing. The authors further indicated that this should be done without discouraging economic activity and without deviating too much from tax systems in other countries. However, establishing an efficient tax system has been a great challenge for most developing countries. This is partly ascribed to the fact that most workers in developing countries are typically employed in agriculture or in small, informal enterprises [3]. Correspondingly, Ghana is predominantly inhabited by people who are self-employed and engaged in agriculture and informal activities. Statistics indicates that 80% of the Ghanaian workforce is employed in the informal sector [4]. The sector is characterized by underemployment, bad working conditions, uncertain work relationships and low wages. The majority of people are living with high income insecurity [4].

Moreover, tax evasion is a well known attitude among workers in Ghana. Annan, et al. [5] found in their study that per capita income, average tax rate, age and inflation are the short run determinants of tax evasion in the country. Consequently, Ghana has not been able to raise sufficient revenue from tax to support financial obligations of past and present governments. This has resulted to high levels of government borrowing over the past years and in present times. Motivated by this background, the aim of the present study is to perform comparative analysis of revenue mobilization instruments under the direct tax of Ghana. The intent is to help identify an efficient tax mobilization potential for the country. The study uses time series tax revenue data extracted from the Asokwa district office of Internal Revenue Service (IRS) of Ghana.

2. Materials and Methods

2.1 Data

The study utilized time series data on annual tax revenue from 2005-2012. This data set was extracted from the Asokwa district office of the Internal Revenue Service (IRS) of Ghana. It contains the revenue collected from PAYE, self-employed tax, company tax and miscellaneous taxes each for the years from 2005 to 2012. Data analysis of this study was computational implemented in SAS 9.1.

2.2 Analysis of Variance (ANOVA)

The basis of the statistical analysis of this study is based on two-way ANOVA. This statistical approach is used to compare the means of populations that are classified in two different ways, or the mean responses in an experiment with two factors. Assuming in a two-way classification table, factor A and factor B are row and column factors respectively, the ANOVA model for two-way classified data is given by Equation (1)

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} \begin{cases} i = 1,2,...,a \\ j = 1,2,...,b \end{cases} \tag{1}$$

where μ is the overall mean effect, α_i is the effect of the ith level of the row factor A, β_j is the effect of jth level of column factor B, and ε_{ij} $NID\left(0,\sigma^2\right)$ random error term [6]. NID means normally and independently distributed. Typically in this model we are interested in testing the hypothesis about the equality of the levels of the row factor,

$$H_0 : \alpha_1 = \alpha_2 = ... = \alpha_a = 0$$
$$H_1 : at\,least\,one\,\alpha_i \neq 0$$

and the equality of the levels of the column factor,

$$H_0 : \beta_1 = \beta_2 = ... = \beta_b = 0$$
$$H_1 : at\,least\,one\,\beta_i \neq 0$$

The ANOVA F test provides a test for the above hypothesis by partitioning the total sum of squares $\left(SS_T\right)$ into variability due to the row factor $\left(SS_A\right)$, variability due to the column factor $\left(SS_B\right)$ and the unexplained variability $\left(SS_E\right)$ given by

$$SS_T = SS_A + SS_B + SS_E \tag{2}$$

When using the ANOVA model we assumed that the errors are normally and independently distributed with mean zero and constant but unknown variance σ^2 i.e. $\varepsilon_{ij} NID\left(0,\sigma^2\right)$. The normality assumption was checked by numerical methods. Specifically, the Shapiro-Wilk test, Kolmogorov-Smirnov test, Cramer-von Mises test and Anderson-Darling test were utilized. The Kolmogorov-Smirnov, Anderson-Darling, and Cramer-von Misers tests are based on the empirical distribution function, which is defined as a set of N independent observations $X_1, X_2, ..., X_N$ with a common distribution function $F\left(X\right)$. The Shapiro-Wilk W is the ratio of the best estimator of the variance to the usual corrected sum of squares estimator of the variance [7].

Moreover, to investigate the constant variance assumption numerical methods such as Bartlett's test and Levene's test were used. In this scenario the hypothesis of interest is

$$H_0 : \sigma_1^2 = \sigma_2^2 = ... = \sigma_k^2$$
$$H_1 : \sigma_i^2 = \sigma_j^2 \; for\,at\,least\,one\,pair\,i \neq j.$$

One procedure for testing the above hypothesis is Bartlett's test. The test statistic is given by

$$B_0 = 2.3026\frac{q}{c},$$

where

$$q = (n-k)\log_{10} MS_W - \sum_{I=1}^{k}(n_i - 1)\log_{10} S_i^2$$

and

$$c = 1 + \frac{1}{3(k-1)}\left[\sum_{i=1}^{k}\left(\frac{1}{n_i - 1}\right) - \frac{1}{n-k}\right].$$

We reject H_0 if

$$B_0 > \chi^2_{k-1}(\alpha),$$

where $\chi^2_{k-1}(\alpha)$ is read from the chi-squared table. Moreover, the above hypothesis for equality of residual variance was numerically investigated using Levene's test. This test proceeds by computing

$$d_{ij} = | y_{ij} - m_i |.$$

where m_i is the median of the observations in treatment group , and d_{ij} are the transformed observations of the of the original observations. Performing the ANOVA F test using the transformed observations, d_{ij} yields Levene's test.

A paired comparison investigation was carried as a post hoc analysis for the ANOVA test about the equality of the levels of the row factor. Specifically, the Least Significance Difference (LSD) was utilized. Suppose that following an ANOVA F test the null hypothesis is rejected, we wish to test

$$H_0 : \mu_i = \mu_j \; for\, all\, i \neq j.$$

This could be done by using the t statistic

$$t_0 = \frac{\bar{y}_{i.} - \bar{y}_{j.}}{\sqrt{MS_E\left(1/n_i + 1/n_j\right)}}$$

and comparing it to $t_{n-k}(\alpha/2)$. An equivalent test declares μ_i and μ_j to be significantly different if $|\bar{y}_{i.} - \bar{y}_{j.}| > LSD$, where

$$LSD = t_{n-k}(\alpha/2)\sqrt{MS_E\left(1/n_i + 1/n_j\right)}$$

In this case $\bar{y}_{i.}$ and $\bar{y}_{j.}$ are the respective mean responses the ith and jth level of the row factor A. If the samples are the same, thus $n_i = n_j = n$ then $LSD = t_{n-k}(\alpha/2)\sqrt{2MS_E/n}$.

3. Results

The methods discussed in the previous section have been applied in this section to analyze the tax data gathered from the Asokwa district office of the IRS in Kumasi, Ghana. The income (in GH¢) of four revenue mobilization instruments; PAYE, self-employed tax, company tax and miscellaneous taxes over the period 2005 to 2012 are considered. The data is classified by revenue mobilization instrument (column) with 4 levels and time (row) with 8 levels (Table 1). The levels of time are the years from 2005 to 2012. Also, the levels of revenue mobilization instrument are PAYE, self-employed tax, company tax and miscellaneous taxes.

Table 1: Tax Revenue data classified by revenue mobilization instrument and time

Block (Time in years)	Treatment Factor (Revenue Mobilization Instruments): Amounts are in GH¢				
	PAYE	Self employed tax	Company tax	Miscellaneous	Block total
2005	919702.32	567264.60	756592.03	347278.86	2590837.81
2006	1418503.74	608740.27	1190911.76	211055.77	3429211.54
2007	1807029.14	819195.60	1171273.90	227634.16	4025132.80
2008	2233732.08	1053694.46	1035331.63	288616.88	4611375.05
2009	3305118.94	1610278.07	1996276.31	173336.06	7085009.38
2010	3967637.75	2336428.37	2145544.80	302391.78	8752002.70
2011	5074150.31	2555997.23	2448881.04	1853652.98	11932681.56
2012	6107448.17	3145571.67	4205569.52	1113334.41	14571923.77
Treatment Total	24833322.45	12697170.27	14950380.99	4517300.90	56998174.61

Figure 1 displays a box plot for the revenue generated from PAYE, self-employed tax, company tax and miscellaneous taxes over the period of 8 years. From this figure, the dots indicate the mean amount of revenue generated from the various revenue mobilization instruments with estimated values of GH¢3,104,165, GH¢1,587,146, GH¢1,868,798 and GH¢564,663 for PAYE, self-employed tax, company tax and miscellaneous taxes, respectively. However, judging by the plots, there appears to be a difference in the mean amounts of revenue generated from PAYE and the remaining three tax types. However, the difference is less pronounced between self-employed tax and company tax.

The results of two-way ANOVA showing the mean difference of the revenue mobilization instruments and time is presented in Table 2. This table provides a test of hypothesis for revenue mobilization instrument and time as stated below:

Hypothesis for Revenue Mobilization Instrument:

H_0 : There is no difference in the mean amount of revenue generated from the revenue mobilization instruments.

H_1 : At least two of the revenue mobilization instruments differ significantly in terms of the amount of income generated.

Hypothesis for Time (in years)

H_0 : The mean amount of revenue generated over time (years) is the same.

H_1 : At least two time periods differ significantly in terms of revenue generated.

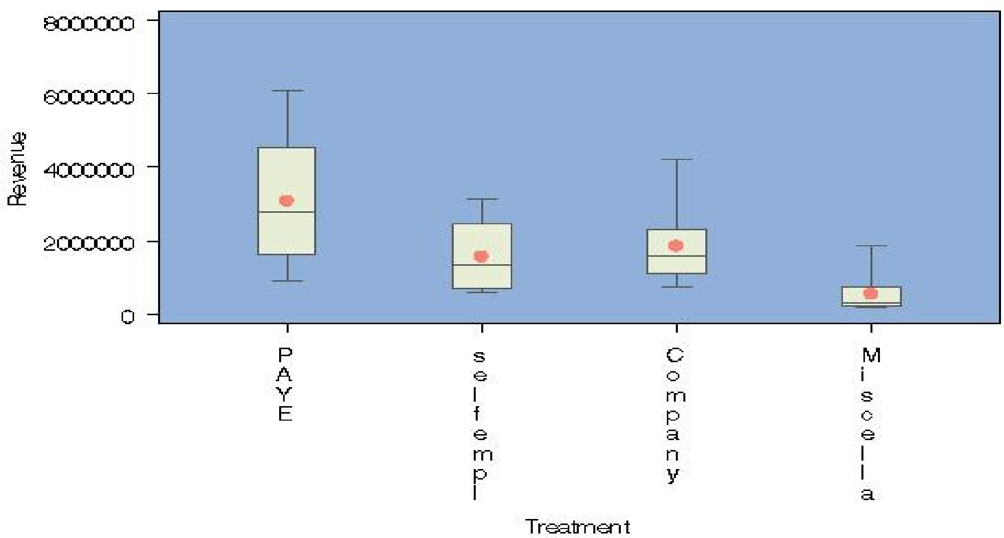

Figure 1: Box Plot for the Mean Income of Different Revenue Mobilization Instruments

From Table 2 the variability in the amount of income due to the four revenue mobilization instruments is 2.620×10^{13} with corresponding p-value less than 0.0001. The p-value leads to the rejection of the null hypothesis, therefore resulting to the conclusion that at least two of the revenue mobilization instruments differ significantly in terms of revenue generated. In addition, the variability in the amount of revenue due to time is 3.284×10^{13} with corresponding p-value less than 0.0001. The p-value indicates that the time factor is significant at 5% level, therefore suggesting that at least two of the time periods differ significantly in terms of revenue generated. The variability in revenue which is unexplained is 9.014×10^{12}.

Table 2: Analysis of Variance for the Tax Revenue Data

Source	DF	Sum of Squares	Mean Square	F Value	P-value
Revenue Mobilization Instrument	3	2.620×10^{13}	8.735×10^{12}	20.35	< 0.0001
Time (Years)	7	3.284×10^{13}	4.692×10^{12}	10.93	< 0.0001
Error	21	9.014×10^{12}	4.292×10^{11}		
Total	31	6.806×10^{13}			

Table 3 provides a numerical test of normality for the ANOVA model. The following hypothesis is considered for the test.

Hypothesis for testing the Normality assumption for the ANOVA

H_0 : The residuals follow the normal distribution.

H_1 : The residuals do not follow the normal distribution.

The tests, thus Shapiro-Wilks test, Kolmogorov-Smirnov test, Cramer-von Mises test and Anderson-Darling tests with p-value of 0.936, 0.150, 0.250 and 0.250 respectively are statistically insignificant at 5% level. This therefore leads to the non-rejection of the null hypothesis, hence indicating that the normality assumption holds.

The numerical test for the constant variance assumption of the two-way ANOVA is presented in Table 4. The hypothesis for the test is stated as follows:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2$$
$$H_1 : \sigma_i^2 = \sigma_j^2 \textit{ for at least one pair } i \neq j.$$

Table 3: Test of Normality for the two-way ANOVA

Test	Statistic	P-Value
Shapiro-Wilk	0.986	0.936
Kolmogorov-Smirnov	0.099	0.150
Cramer-von Mises	0.043	0.250
Anderson-Darling	0.238	0.250

Table 4: Test of Homogeneity of Revenue Variance for the two-way ANOVA

Tests	Bartlett's Test	Levene's Test
Chi Squared value	2.94	
F-value		0.33
P-value	0.4073	0.3737

From Table 4 the Bartlett's test with chi squared value of 2.94 and corresponding p-value of 0.4073 is insignificant at 5% level. Similarly, the Levene's test is not significant (F-value=0.33; p-value=0.3737). These tests lead to the non-rejection of the null hypothesis, therefore giving a strong indication that the constant variance assumption holds. Table 5 displays a post hoc test based on the Least Significant Difference (LSD) approach for the ANOVA model.

From the table there is a significant difference between PAYE and self-employed tax in that their absolute mean difference (AMD) of 1,235,367 is greater than the LSD which is 681,251. Similarly, PAYE and company tax, and PAYE and miscellaneous tax differ significantly. Self-employed tax and miscellaneous taxes, and company tax and miscellaneous taxes also differ significantly. However, there is no significant difference between self-employed tax and company tax because their AMD of 281,652 is less than the LSD of 681,251.

Table 5: Comparing LSD and Absolute Mean Amount of Revenue Difference

Paired	Absolute Mean Difference of Revenue	Rank	Remarks
PAYE and Self-employed tax	1,235,367	4	Differ significantly
PAYE and Company tax	1,517,019	2	Differ significantly
PAYE and Miscellaneous tax	2,539,502	1	Differ significantly
Self-employed and Company	281,652	6	No difference
Self-employed and Miscellaneous	1,022,483	5	Differ significantly
Company and Miscellaneous tax	1,304,135	3	Differ significantly
LSD	681,251		

Moreover, in a decreasing order of magnitude it is evident that much of the variability resulted from the difference between PAYE and miscellaneous tax (GH¢ 2,539,502), PAYE and company tax (GH¢ 1,517,019), company and miscellaneous tax (GH¢ 1,304,135), PAYE and self-employed tax (GH¢ 1,235,367), self-employed and miscellaneous tax (GH¢ 1,022,483), and self-employed and company tax (GH¢ 281,652).

4. Conclusions

This study has investigated the difference in amounts of revenue generated from PAYE, self-employed tax, company tax and miscellaneous taxes of the Asokwa district office of IRS. With the mean amount of revenue from PAYE, self-employed tax, company tax and miscellaneous tax of GH¢3,104,165, GH¢1,587,146, GH¢1,868,798 and GH¢564,663 respectively, the results of the two-way ANOVA depicted a significant difference between at least two of the revenue mobilization instruments. Specifically, results from a post hoc analysis revealed that PAYE and self-employed tax, PAYE and company tax, PAYE and miscellaneous taxes, self-employed and miscellaneous taxes, and company tax and miscellaneous taxes differ significantly in terms of revenue generated. However, no significant difference was found between self-employed tax and company tax. It is understandable that the revenue from self-employed tax which comes from the largest workforce is equated to revenue from the few companies in the district. In terms of revenue mobilization, the larger size of the informal sector appears to be a potential for the district and for that matter Ghana at large. There is therefore the need to implement policies that can boost this sector. By such means, Ghana can establish an efficient tax system with utmost focus on the informal sector.

References

1. Ali-Nakyea, A. (2008). *Taxation in Ghana, Principles, Practices and Planning* (2nd ed.), Accra: Blackmask Ltd.
2. Price Water House. 2008: *Tax Facts and Figures: A Quick Guide to Taxation in Ghana*
3. http://www.pwc.com/en_GH/gh/pdf/tax-facts-and-figures-2008.pdf (Accessed: November, 20, 2014).
4. Tanzi, V. and Zee, H. 2001: *Tax Policy for Developing Countries. International Monetary Fund.* http://www.imf.org/external/pubs/ft/issues/issues27/ (accessed: October, 30, 2014).
5. Osei-Boateng, C. and Ampratwum E. 2011: *The Informal Sector in Ghana.*
6. http://www.fesghana.org/uploads/PDF/FES_InformalSector_2011_FINAL.pdf (accessed: November, 16, 2014).
7. Annan, B., Bekoe W., and Nketiah-Amponsah E. 2014: Determinants of Tax Evasion in Ghana: 1970-2010. *International Journal of Economic Sciences and Applied Research* 6 (3): 97-121.
8. Montgomery, D.C. 2001: *Design and Analysis of Experiments*. 5th edition. John Wiley & Sons, Inc.
9. Shapiro, S.S, and M. B. Wilk, 1965: An Analysis of Variance Test for Normality (Complete Samples). *Biometrika,* 52(3/4) :591-611.

DVIFAKTORINĖ DISPERSINĖ ANALIZĖ PAJAMŲ SURINKIMO SKIRTUMAMS TIRTI

Kwaku F. Darkwah, Richard Tawiah, Maxwell Adu-Gyamfi

Santrauka. Ganos neformalusis sektorius pasižymi nepakankamu užimtumu, blogomis darbo sąlygomis, neužtikrintais darbo santykiais ir mažu uždarbiu. Šiame sektoriuje užimta apie 80 % darbo jėgos. Todėl efektyvios mokesčių sistemos sukūrimas, kuri galėtų surinkti pakankamai biudžeto pajamų pagrindinėms išlaidoms finansuoti išvengiant valstybės skolinimosi, yra didelis iššūkis. Naudojant Ganos biudžeto įplaukų surinkimo laiko eilutės duomenis, šiame straipsnyje nagrinėjami keturių mokesčių surinkimo instrumentų skirtumai, būtent sau dirbančiųjų mokesčiai, įmonių mokesčiai, pajamų mokesčiai ir kiti mokesčiai. Straipsnyje taikoma dvifaktorinė dispersinė analizė kaip pagrindinis metodas mokesčių skirtumams tirti. Teorinio modelio tinkamumas tiriamas skaitiniais metodais. Skirtingų surinkimo instrumentų surinktos įplaukos labai skiriasi, išskyrus sau dirbančiųjų mokesčius ir įmonių mokesčius. Įmonių, sau dirbančių ir kiti mokesčiai buvo maži, palyginti su pajamų mokesčiais. Šiuo tyrimu atkreipiamas dėmesys į poreikį stiprinti neformalųjį sektorių, siekiant pagerinti sau dirbančiųjų mokesčių surinkimą.

Reikšminiai žodžiai: dispersinė analizė, Gana, mokesčių surinkimo instrumentas, apmokestinimas.