A Bayesian Model for Estimation of Population Proportions

A paper presented at JSM 2011 held at the Convention Center, Miami Beach, Florida, USA.

By

Okafor, R. O. and Mbata, U. A.

Department of Mathematics, Faculty of Science, University of Lagos, Akoka, Lagos, Nigeria.

okaforray@yahoo.com, mbataugochukwu@yahoo.com

ABSTRACT

- In recent times and even up to now, traffic congestion and parking difficulties, especially during morning and evening rush hours, have become a major concern to members of the University of Lagos (UNILAG) community and visitors alike. UNILAG has witnessed unprecedented growth in student's enrolment, in the last ten years or so culminating in the current total enrolment of more than thirty-five thousand students, of which about twenty-five thousand are undergraduates. In order to study the worrisome traffic situation at UNILAG in the wake of these large numbers, independent, though similar, sample surveys of undergraduate students of the eight faculties on the main campus of UNILAG were conducted in 2007. The purpose of the surveys was to collect data on undergraduate students who owned or used motor vehicles on campus. Furthermore, to investigate possible temporal trends, the surveys were repeated in 2009. The types of data obtained from the surveys provided veritable impetus for the application of Empirical Bayes (EB) analysis to estimate the proportions of students of individual faculties who used motor vehicles on campus roads during the periods under reference. The EB technique, being a Bayesian method, combines prior information and sample information in a manner that "shrinks" an EB estimator towards the sample estimator if a vague prior (proper prior with a large variance) is used. The main result from this research is that in 2007 about one in four students used motor vehicles, and this result held almost across the eight faculties. Although results of the 2009 surveys were generally similar there were faculties that recorded some reduction in the estimated proportions of students who used motor vehicles.

- **Key Words:** Estimation, Empirical Bayes, Faculties at Main Campus, Proportions using Vehicles, Sample Surveys, Traffic Congestion, Undergraduates, UNILAG, Use of Motor Vehicles.

INTRODUCTION

- For sometimes now vehicular traffic congestion has become a major problem to commuters using the main access roads on the main campus (at Akoka) of the University of Lagos (hereafter UNILAG), Lagos, Nigeria.
- In 2007, a group of eight final-year students of Statistics of the Department of Mathematics, UNILAG, were selected to conduct parallel sample surveys across the eight Faculties at Akoka; one student being assigned to a Faculty.
- The surveys formed part of the said students' final-year projects.
- The purpose of the surveys was to collect data on undergraduate students (from each Faculty) who used motor vehicles on campus roads.
- The main statistical task was to estimate the proportion of students from each Faculty who used motor vehicles during the session 2006/2007.
- For comparison purposes, the surveys were repeated in 2008/2009 session by another set of final-year statistics students of the Department of Mathematics.
- This paper presents a re-analysis of the data collected from the surveys.

EMPIRICAL BAYES METHOD

- The method of Empirical Bayes (EB) continues to receive increasing popularity since its introduction about four decades ago by Robbins (1955) .
- Basically, Empirical Bayes (EB) method is based on Bayes Theorem, and is applied to estimation problems in which the statistician has the good fortune that supplementary data (on related problems) are available.
- The surveys under reference yielded the types of data for which the EB method is appropriate.
- EB was adopted in this research because it has the capacity to combine estimates of independent but related studies, and hence improve the results of estimation.
- Some notable works include Rubin (1980), Deely & Lindley (1981), Hui & Berger (1983), Morris (1983), Wong (1987), Smith (1991), Raghunathan (1993), Altman & Casella (1995), Efron (1996) and Okafor (1999).
- More recent works and applications include Link & Hahn (1996), Carlin & Louis (2000a, 2000b), Natarajan & Kass (2000), Brandel (2004), Lee (2004), Theobald & Wuttke (2006), Ogundeji & Okafor (2010), and Okafor et al (2010).

BETA-BINOMIAL CONJUGATE MODEL

- The EB model adopted in this paper is for the estimation of the proportion of students who used motor vehicles in each Faculty.
- The Beta – Binomial model is defined as follows: Let

  n_i = number of registered students in ith Faculty, $i = 1.2,\ldots,k$ where $k = 8$ is number of Faculties.

  $$X_{ij} = \begin{cases} 1, & \text{if jth student in ith Faculty used motor vehicle} \\ 0, & \text{otherwise} \end{cases}$$

  $Y_i = \sum_{j=1}^{n_i} X_{ij}$ is the total number of students who used motor vehicles in ith Faculty.

- We specify $Y_i \sim Binomial\ (n_i, P_i)$ and in the Bayesian framework, we model $P_i \sim Beta\ (\alpha, \beta]$, where α, β are unknown parameters which would be estimated in accordance with the principle of Empirical Bayes.
- Under the Beta – Binomial conjugate modeling, the EB estimator $\hat{P}_{EB}$ of P_i is the weighted average:

  $$\hat{P}_{EB} = \hat{\lambda}\hat{P}_o + (1 - \hat{\lambda})\hat{P}_i$$

  where $\hat{P}_o$ is the estimated mean of the prior density and $\hat{P}_i$ is the maximum likelihood estimate of P_i
- Thus,

  $$\hat{P}_o = \frac{\hat{\alpha}}{\hat{\alpha} + \hat{\beta}}, \hat{P}_i = \frac{Y_i}{n_i}, \quad \hat{\lambda} = \frac{\hat{\alpha} + \hat{\beta}}{\hat{\alpha} + \hat{\beta} + n_i}$$

- Note that $\hat{\alpha}$ and $\hat{\beta}$ are obtained by the method of maximum likelihood using some enabling re-parameterizations .

ANALYSIS OF 2007 DATA

Table 1a: Sample Estimates of Proportions and Estimated Variances

Faculty	N_i	n_i	Y_i	$\hat{P}_i$	$Var(\hat{P}_i)$	$Se(\hat{P}_i)$	LCL	UCL	Width
Arts	2313	231	63	0.2727	0.000859	0.029303	0.2153	0.3302	0.11
Bus. Admin.	2959	231	51	0.2208	0.000745	0.027290	0.1673	0.2743	0.11
Education	2583	225	69	0.3067	0.000945	0.030741	0.2464	0.3669	0.12
Engineering	3554	236	41	0.1737	0.000608	0.024663	0.1254	0.2221	0.10
Environ. Sc.	1642	222	59	0.2658	0.000879	0.029648	0.2077	0.3239	0.12
Law	900	254	106	0.4173	0.000957	0.030941	0.3567	0.4780	0.12
Science	2859	211	60	0.2844	0.000964	0.031056	0.2235	0.3452	0.12
Social Sc.	2291	222	62	0.2793	0.000907	0.030111	0.2203	0.3383	0.12
UNILAG	19101	1832	511	0.2789	0.000110	0.010478	0.2584	0.2995	0.04

Table 1b: Empirical Bayes Estimates of Proportions and Estimated Variances

Faculty	$\hat{\alpha}$	$\hat{\beta}$	$\widetilde{P}_{EB}$	$Var(\widetilde{P}_{EB})$	$Se(\widetilde{P}_{EB})$	LCL	UCL	Width
Arts	75.38	200	0.2737	0.000719	0.026820	0.2297	0.3177	0.09
Bus. Admin.	63.38	212	0.2302	0.000641	0.025319	0.1886	0.2717	0.08
Education	81.38	188	0.3021	0.000780	0.027924	0.2563	0.3479	0.09
Engineering	53.38	227	0.1904	0.000548	0.023405	0.1520	0.2288	0.08
Environ. Sc.	71.38	195	0.2680	0.000734	0.027085	0.2235	0.3124	0.09
Law	118.38	180	0.3967	0.000799	0.028274	0.3504	0.4431	0.09
Science	72.38	183	0.2834	0.000792	0.028145	0.2373	0.3296	0.09
Social Sc.	74.38	192	0.2792	0.000753	0.027435	0.2342	0.3242	0.09
UNILAG	523.38	1353	0.2789	0.000107	0.010350	0.2620	0.2959	0.03

ANALYSIS OF 2009 DATA

Table 2a: Sample Estimates of Proportions and Estimated Variances

Faculty	N_i	n_i	Y_i	$\hat{P}_i$	$Var(\hat{P}_i)$	$Se(\hat{P}_i)$	LCL	UCL	Width
Arts	1404	192	53	0.2760	0.001041	0.032262	0.2128	0.3393	0.13
Bus. Admin.	1915	248	73	0.2944	0.000838	0.028940	0.2376	0.3511	0.11
Education	2651	331	43	0.1299	0.000341	0.018479	0.0937	0.1661	0.07
Engineering	1851	204	54	0.2647	0.000954	0.030889	0.2042	0.3252	0.12
Environ. Sc.	1091	188	25	0.1330	0.000613	0.024764	0.0844	0.1815	0.10
Law	882	100	12	0.1200	0.001056	0.032496	0.0563	0.1837	0.13
Science	2089	182	34	0.1868	0.000835	0.028891	0.1302	0.2434	0.11
Social Sc.	1823	200	56	0.2800	0.001008	0.031749	0.2178	0.3422	0.12
UNILAG	13706	1645	350	0.2128	0.000102	0.010091	0.1930	0.2325	0.04

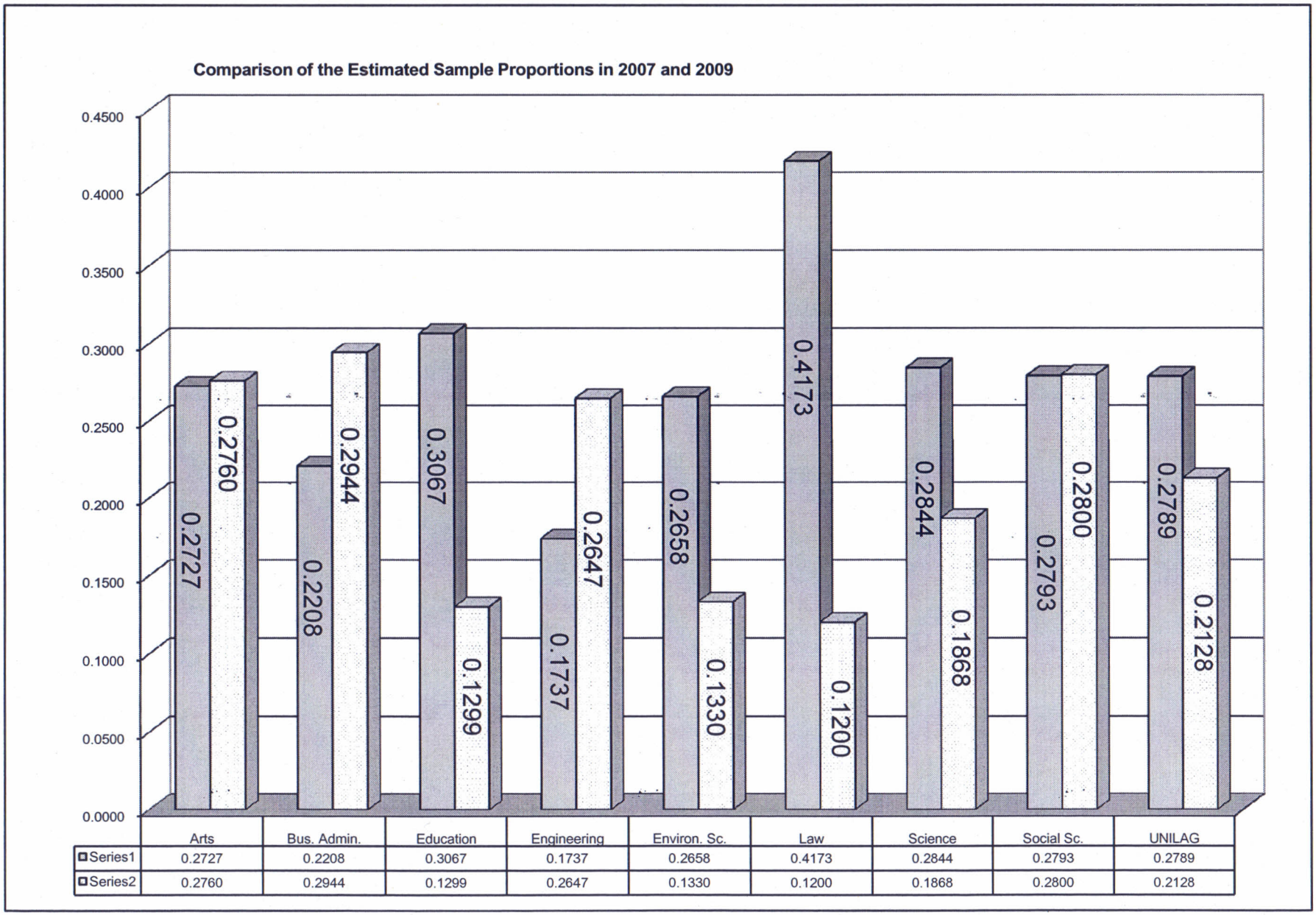
Table 2b: Empirical Bayes Estimates of Proportions and Estimated Variances

Faculty	$\hat{\alpha}$	$\hat{\beta}$	$\tilde{P}_{EB}$	$Var(\tilde{P}_{EB})$	$Se(\tilde{P}_{EB})$	LCL	UCL	Width
Arts	60.04	165.06	0.2667	0.000865	0.029411	0.2185	0.3150	0.10
Bus. Admin.	80.04	201.06	0.2847	0.000722	0.026869	0.2407	0.3288	0.09
Education	50.04	314.06	0.1374	0.000325	0.018020	0.1079	0.1670	0.06
Engineering	61.04	176.06	0.2575	0.000803	0.028335	0.2110	0.3039	0.09
Environ. Sc.	32.04	189.06	0.1449	0.000558	0.023621	0.1062	0.1837	0.08
Law	19.04	114.06	0.1431	0.000914	0.030236	0.0935	0.1927	0.10
Science	41.04	174.06	0.1908	0.000714	0.026729	0.1470	0.2346	0.09
Social Sc.	63.04	170.06	0.2705	0.000843	0.029031	0.2228	0.3181	0.10
UNILAG	357.04	1321.06	0.2128	0.000100	0.009988	0.1964	0.2291	0.03

Table 3: Comparison of the Estimated Proportions in 2007 and 2009

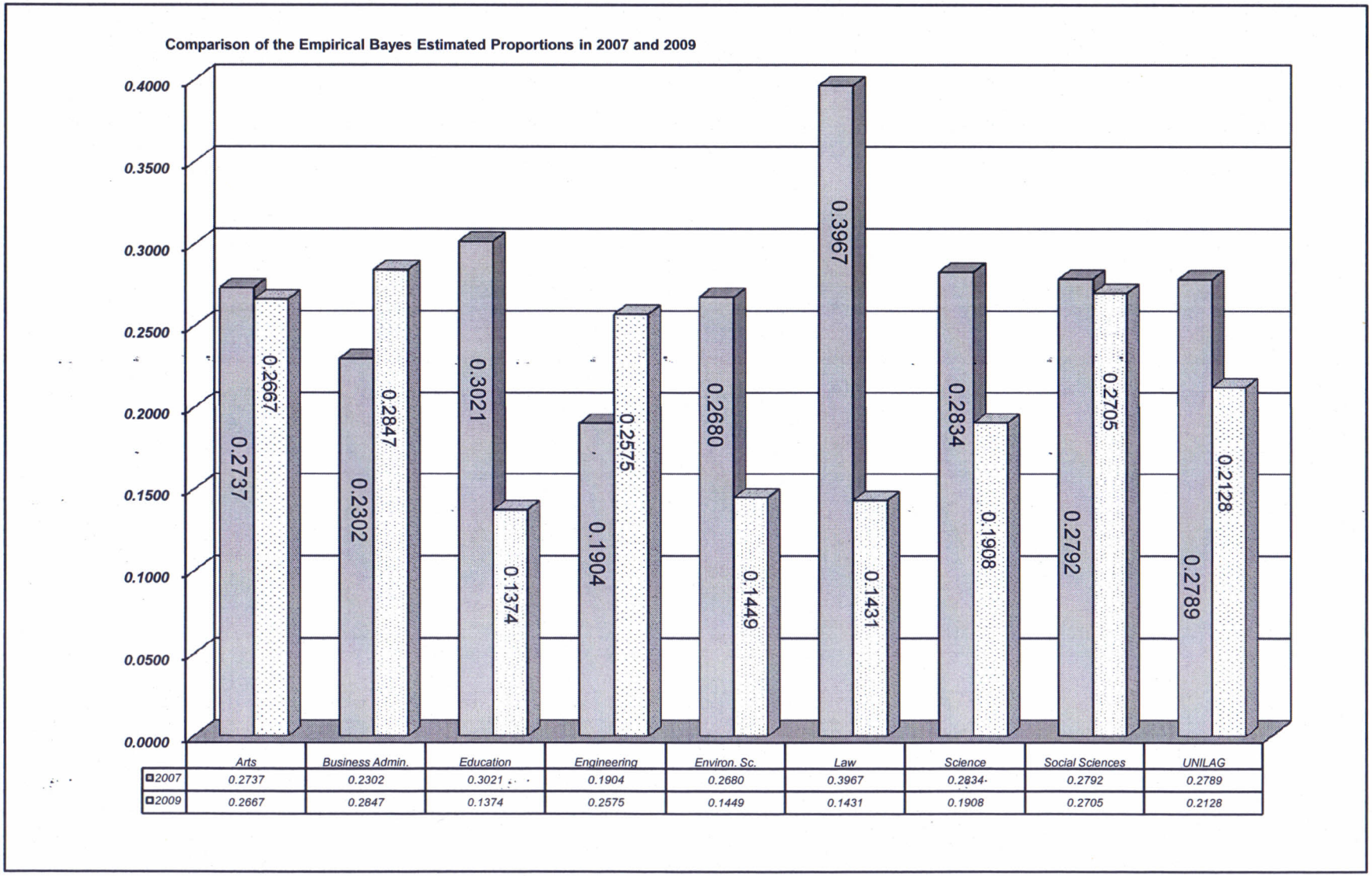
Faculty	$\tilde{P}_{EB}$ (2007)	$\hat{P}_i$ (2007)	$\tilde{P}_{EB}$ (2009)	$\hat{P}_i$ (2009)
Arts	0.2737	0.2727	0.2667	0.2760
Business Admin.	0.2302	0.2208	0.2847	0.2944
Education	0.3021	0.3067	0.1374	0.1299
Engineering	0.1904	0.1737	0.2575	0.2647
Environmental Sc.	0.2680	0.2658	0.1449	0.1330
Law	0.3967	0.4173	0.1431	0.1200
Science	0.2834	0.2844	0.1908	0.1868
Social Sciences	0.2792	0.2793	0.2705	0.2800
UNILAG	0.2789	0.2789	0.2128	0.2128

Figure 1: BAR CHART



	Arts	Bus. Admin.	Education	Engineering	Environ. Sc.	Law	Science	Social Sc.	UNILAG
Series1	0.2727	0.2208	0.3067	0.1737	0.2658	0.4173	0.2844	0.2793	0.2789
Series2	0.2760	0.2944	0.1299	0.2647	0.1330	0.1200	0.1868	0.2800	0.2128

Figure 2: BAR CHART



	Arts	Business Admin.	Education	Engineering	Environ. Sc.	Law	Science	Social Sciences	UNILAG
2007	0.2737	0.2302	0.3021	0.1904	0.2680	0.3967	0.2834	0.2792	0.2789
2009	0.2667	0.2847	0.1374	0.2575	0.1449	0.1431	0.1908	0.2705	0.2128

DISCUSSION OF RESULTS

- The analyses have shown that the overall proportion of university undergraduate students who used personal motor vehicles on campus was 0.2789 (in 2007) and 0.2128 (in 2009). This implies that nearly one out of four undergraduate students use motor vehicle on Unilag main campus during the two-year period.
- Generally, the estimated proportions generated by the EB process have shown clearly that there was a remarkable decline in the use of motor vehicles by undergraduate students in 2009 compared to the results in 2007.
- This reduction may be attributed to some economic changes experienced from 2007 to 2009 such as collapse of the stock market and global financial meltdown.
- Also, the reduction may be because of the recent introduction of paid car parking lots by the university authorities, which probably reduced the propensity of students to bring motor vehicles to campus.

REFERENCES

[1] Altman, N. S. & Casella, G. (1995): Nonparametric Empirical Bayes Growth Curve Analysis. Journal of the American Statistical Association, Vol. 90, pp. 508-515.

[2] Brandel, J. (2004): Empirical Bayes Methods for missing data analysis. Department of Mathematics Uppsala University Project Report.

[3] Carlin, B. P. & Louis, T. A. (2000): Empirical Bayes: Past, Present and Future. Journal of the American Statistical Association, Vol. 95, No. 452, pp.1286-1289.

[4] Carlin, B. P. & Louis, T. A. (2000): Bayes and Empirical Bayes Methods for Data Analysis. Boca Raton, Florida: Chapman and Hall/CRC Press.

[5] Deely, J. & Lindley, V. (1981): Bayes Empirical Bayes. Journal of the American Statistical Association, 76, pp. 833-841.

[6] Efron, B. (1996): Empirical Bayes Methods for Combining Likelihoods (with discussion). Journal of the American Statistical Association, Vol. 91, pp. 538-565.

[7] Hui, S. & Berger, J. (1983): Empirical Bayes estimation of rates in longitudinal studies. Journal of the American Statistical Association, 78, pp. 753-760.

[8] Lee, P. (2004): Bayesian Statistics An Introduction. New York.: Hodder Arnold.

[9] Link, W. A. & Hahn, D. C. (1996): Empirical Bayes Estimation of Proportions with Application to Cowbird Parasitism Rates. Journal of Ecological Society of America, 77(8), pp. 2528-2537.

[10] Smith, P. J. (1991): Bayesian analyses for a multiple capture-recapture model. Biometrika 78, pp. 399-407.

REFERENCES (CONTD.)

[11] Morris, C. (1983): Parametric empirical Bayes inference: Theory and applications. Journal of the American Statistical Association, Vol. 78, pp. 47-65.

[12] Ogundeji, R. K. & Okafor, R. O. (2010): Empirical Bayes Approach to Estimate Mean CGPA of University Graduates. International Journal of Applied Mathematics & Statistics Vol. 17; No. J10; pp. 77-89.

[13] Okafor, R. (1999): Using an Empirical Bayes Model to estimate currency exchange rate. Journal of Applied Statistics, Vol. 26, No. 8, 973-983.

[14] Okafor R.O., Opara A.I., Ogundeji R.K., Mbata U.A., & Olalude G. A. (2010): Empirical Bayes Model for Inference on VehicularTraffic Density at the Main Campus of University of Lagos, Lagos, Nigeria. International Journal of Statistics and Systems, Vol. 5 No. 1, pp. 53–62.

[15] Robbins, H. (1955): An empirical Bayes approach to Statistics. Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1.

[16] Rubin, D. (1980): Using empirical Bayes techniques in the law school validity studies. Journal of the American Statistical Association, 75, pp. 801-816.

[17] Theobald & Wuttke (2006): Empirical Bayes hierarchical models for regularizing maximum likelihood estimation in the matrix Gaussian Procrustes problem. The National Academy of Sciences of the USA, vol. 103 no. 49, pp. 18521-18527.

[18] Wong, G. (1987): Bayesian models for directed graphs. Journal of the American Statistical Association, 82, pp. 140-148.

[19] Natarajan, R. & Kass, R. E. (2000): Reference Bayesian methods for Generalized Linear mixed models. Journal of the American Statistical Association, 95, pp. 227-237.

[20] Raghunathan, T. E. (1993): A Quasi-Empirical Bayes methods for Small Area Estimation. Journal of the American Statistical Association, 88, pp. 1444-1448.