International Journal of Statistics and Applications 2018, 8(6): 316-322
DOI: 10.5923/j.statistics.20180806.05

Modeling Patients' Choice of Healthcare Providers: The Fractional Data Approach

Dimitris Zavras

Department of Health Economics, National School of Public Health, Athens, Greece

Abstract According to the fractional data approach, the utilization of public or/and private healthcare services constitutes a fraction of total health services use, with values in the unit interval. The objective of this paper is to present the fractional data approach of health care provider choice and to check the suitability of alternative models that could be used to analyse such fractions. Data from a Panhellenic cross-sectional survey carried out in 2006 in a random, stratified sample of 4003 adults was used to compare the appropriateness and suitability of the alternative models. The dependent variable in the analysis was the "the proportion of both private and public health services utilization" and the study focused on users that utilized healthcare services more than once. Quasi Maximum Likelihood Model was compared with the use of Ordinary Least Squares Model. The results of the analyses indicate that the Ordinary Least Squares Model gives a satisfactory fit to data, while its marginal effects are similar to those of the Quasi Maximum Likelihood Model. The results of the study confirm existing international literature, according to which, in the case of proportions, Ordinary Least Squares Models can provide satisfactory results.

Keywords Choice of Healthcare Providers, Fractional Data Approach, Utilization of Health Services, Ordinary Least Squares Model, Quasi Maximum Likelihood Model

1. Introduction

Healthcare seeking behavior is one of the main topics studied in health economics [1] and health services research [2]; in particular, understanding how patients make decisions on the type of service to utilize has a central role in understanding health system functioning.

Utilization of health services occurs when the individual acts on a health need to demand and receive health services [3]. Starting in the 60s, the difference between medical need, demand and utilization has been intensely studied and predictors of each of the three parameters have been identified. With advances in science and technology scientists now are able to explore not only individual characteristics (income, education, occupation, medical insurance, age, gender, family structure, exposures to various environmental risk factors) that affect need, demand or actual utilization, but also intrinsic (i.e. genetic or epigenetic effects) or external predictors such as social structure, small area geographic influences and more.

Of paramount importance is to understand the quantitative aspect of the studies and the statistical methods used to analyze the data. How data is analyzed is important to warranty an unbiased, valid result. Because each statistical method requires that certain assumptions hold, understanding the assumptions and the effects of their violation offers a critical skill to researchers and meta-analysts to interpret the published literature correctly.

The majority of studies on patient's selection of healthcare provider are based on discrete choices [4] represent classes of categorical variables. The models used (Logit or Probit) are based on the theoretical linkage between the economic theory of consumer choice and the mathematical modeling of discrete analysis [5, 6].

The theoretical framework behind the discrete choice approach for studying how consumers make healthcare choices was developed primarily by Gertler et al. [7] and Cameron et al. [8]; people in need of a physician make a discrete selection from a variety of alternatives, based on a comparison of the utility of each alternative, which depends both on the features of the users and on the features of the healthcare services [9-11].

According to the discrete choice approach, studying how groups of patients select healthcare providers (patients as consumers of services and decision makers) requires capturing and defining the following elements:

a) the actual choices available, and sets of alternatives, known as *the consideration set* (the consideration set is the dependent variable of the analysis);

* Corresponding author:
dzavras@esdy.edu.gr (Dimitris Zavras)
Published online at http://journal.sapub.org/statistics

b) the attributes or characteristics of each alternative choice;

c) the attributes or characteristics of the decision makers, and how these interface with the attributes of the alternative choices (via a rule for combining them);

d) a model of individual choice and behavior, as well as the distribution of behavior patterns in the population.

The understanding of how choices are made depends on the nature and availability of the data surrounding the above attributes [12]. In other words, the appropriate specification of the consideration set depends on the objective of the survey and the data availability [13]. Although access to healthcare is an important determinant of healthcare utilization, it also serves as a determinant of the consideration set itself. The phenomenon of access to healthcare from the perspective of how easy it was for users to enter into the health system was analytically and in depth examined by Penchansky and Thomas associating access to healthcare with *availability, accessibility, accommodation, affordability and acceptability* [14]. It must be noted that availability, out of the five *As*, is directly associated with the determination of the choice set.

When determining the consideration set, important variables are the types of healthcare used, and their frequency. The most common frequencies reported in the literature related to the number of visits to alternative healthcare providers refer to time spans ranging from 1 month to 1 year [15, 16].

In theory, the consideration set needs to present three characteristics: i) the alternatives must be mutually exclusive, ii) the choice set must be exhaustive and iii) the number of alternatives must be finite [17]. However in reality, since the first and second criteria are not restrictive, there is a possibility of: a) observing combinations between the alternatives and; b) no-selection can be included in the consideration set.

Essentially, starting from a quantitative discrete dependent variable (number of visits), it is often possible to transform utilization to a discrete qualitative variable in order to capture and analyze the use of all alternatives within the consideration set. One way to code choice is presented in Table 1.

Table 1. Utilization of Alternative Health Services

Number of Visits in Private Practice	Number of Visits in Public Services	Alternatives
0	0	1: No use
a>0	0	2: Use of Private Practice Only
0	b>0	3: Use of Public Services Only
a>0	b>0	4: Combinations

This coding scheme of the alternatives could be considered as incorporating ordering or not.

In the case of health systems where private and public sectors co-exist as in Greece, the alternatives are: a) use of public sector healthcare services, b) use of private sector healthcare services and c) combined use of healthcare services of public and private sector.

However, little attention has been given until now to the study of the combined use of healthcare services, something that may be due to the short recall period [18]. In studies with short observation period and small number of visits to healthcare facilities, the potential of combinations seems to be limited due to the fact that the opportunities for combined or mixed use may not arise in such limited time interval. Recall period is of critical importance; as the time interval becomes longer, the numbers of possible alternatives are likely to increase [19]. Respectively, for short observation periods, the consideration set may not include combination of alternatives.

The objective of this paper is to present the fractional data approach of health care provider choice and to check the suitability of alternative models that could be used to analyse such fractions.

2. Material and Methods

As mentioned above, healthcare utilization is usually expressed as a non-negative number, and more specifically as a number of visits to a given provider in a given time-period, with time usually varying from 1 to 12 months depending on the aims of the study and the nature of the medical need that is addressed in the study [15, 16].

The fractional data approach takes into account the fact that the utilization of health services of the private or/and public sector (healthcare services of alternative sectors) is a fraction of total healthcare use, with values in the unit interval.

The total utilization of health services is calculated as the sum of total utilization of private health services and total utilization of public health services:

$$\text{Total Utilization of Health Services} = \text{Total Utilization of Private Health Services} + \text{Total Utilization of Public Health Services} \quad (1)$$

Following that, the variables of the total utilization of private sector and public sector health services are divided by the total utilization of health services:

$$P_{\text{Total Utilization of Private Health Services}} = \text{Total Utilization of Private Health Services} / \text{Total Utilization of Health Services} \quad (2)$$

$$P_{\text{Total Utilization of Public Health Services}} = \text{Total Utilization of Public Health Services} / \text{Total Utilization of Health Services} \quad (3)$$

The sum of the aforementioned proportions equals to 1.

Values in the (0, 1) interval express the combined utilization of private and public healthcare services, a value of 0 expresses non-utilization of private or public healthcare services and a value of 1 expresses the exclusive utilization of private or public healthcare services. Similarly, a value of 0.7 for utilization of private health services means that 70% of the total healthcare utilization is due to private utilization.

As these proportions describe quantitatively the parts of a

whole, they constitute an example of two-dimensional compositional data.

It is evident that this particular approach focuses solely on the users of healthcare services.

The main advantage of the fractional data approach results from the information contained in the data. As it is known, the scale of measurement defines an information hierarchy [20]. Depending on the scenario, this hierarchy may include the nominal scale, the ordinal scale, the interval scale and the ratio scale. The lowest level in the measurement hierarchy is the nominal scale and the middle level is the ordinal scale [21] i.e. the discrete choices about healthcare provider. Quantitative measures, such as the aforementioned proportions, convey the most information about a variable because they allow the researcher to classify, order, and array values along a clearly delineated common scale or standard [22].

In order to check the suitability of alternative models used in scientific literature, data from a Panhellenic cross-sectional survey from 2006 were used.

The stratified random sampling method was used to select the sample.

Sample selection strata were based on the 2001 Census of the Hellenic Statistical Authority and they were defined by age, gender, urbanity status of permanent residence and prefecture based on Nomenclature of Territorial Units for Statistics NUTS II.

Due to lack of information related to the distribution of health services utilization with the stratification variables, the formula of simple random sampling was used for the sample size calculation [23]:

$$n=p(1-p)z^2/d^2 \quad (4)$$

where z = 1.96 (abscissa of the normal curve that cuts off an area of α (0.05) at the tails), d=1.55% (margin of error), and $p_{users\text{-}non\ users}$ = 0.5 (estimated proportion of the response that returns the maximum sample size). The sample size was found equal to n=3997.

Subjects were asked to report on experiences during the preceding year. The survey used a questionnaire based on World Health Organization methodology [24], which had been validated in the past.

The sample consisted of: a) individuals insured in Social Insurance Organisation (I.K.A) (45.77%); b) individuals insured in other insurance organisations (48.26%) και c) uninsured individuals (4,07%) (a small proportion of respondents did not answer the question about health insurance coverage) [25, 26].

The distinction mentioned above was necessary since the social insurance status determined the health services available (individuals insured in I.K.A could use public health services, private health services and I.K.A primary health centers, which was not the case for individuals insured in other insurance organisations).

The study focused on the group of individuals insured in other insurance organisations (n=1932) and more specifically to those individuals using both public and private health services (n=221) and had total health services utilization more than once (Table 2).

Table 2. Healthcare Provider Choice Alternatives

Alternative	n (%)
No Use	639 (33.07)
Use of Private Practice Only	708 (36.64)
Use of Public Services Only	364 (18.84)
Combinations	221 (11.43)
Total	1932

Data collection involved a personal, face-to-face interview. The questionnaires were pilot tested prior to the main data collection phase (June 2006). The interviews were carried out by specially trained professional interviewers and standard quality assurance procedures were used (reconfirmed 15% of the original interview data).

The dependent variable was the "the proportion of both private and public health services utilization". Since the response variable takes values in the interval of (0, 1) and could be considered univariate [27], I had the option to fit and compare four different types of models: 1) an Ordinary Least Squares (OLS) Model on data, transformed via log-ratio [28]; 2) a Beta Regression (BR) Model which is a special case of the Dirichlet Model [29, 30]; 3) an OLS Model on original data [31] and 4) a Quasi Maximum Likelihood (QML) Model [32].

Potential predictors (independent variables) in the models were the following: a) gender (1: women, 2: men); b) age; c) self-reported health status (1: very bad, 2: bad, 3: moderate, 4: good and 5: very good); d) existence of chronic health condition (1: no, 2: yes); e) education level (1: primary education, 2: secondary education, 3: tertiary education); f) income level (1: no income, 2: 1-500€, 3: 501-1000€, 4: 1001-1500€, 5: 1501-2000€, 6: 2001-3000€ and 7: 3001€+); g) employment status (1: working, 2: unemployed, 3: retiree, 4: housekeeper, 5: student or soldier, 6: other); h) public health insurance (1: no, 2: yes); i) private health insurance (1: no, 2: yes); j) urbanity status of permanent residence (1: rural, 2: urban); k) geographic prefecture (1: Attica, 2: East Macedonia and Thrace, 3: West Macedonia, 4: Central Macedonia, 5: Epirus, 6: Thessaly, 7: West Greece, 8: Central Greece, 9: Islands of Northern Aegean, 10: Islands of Southern Aegean, 11: Peloponnese, 12: Ionian Islands, 13: Crete), l) total utilization of healthcare services [33] (1: 2-3, 2: 4-6, 3: 7-10, 4: 11-15, 5: 16+).

From the variables mentioned above, age and ordinal variables (self-reported health status, education, income and total utilization of healthcare services) were treated as continuous.

From each nominal variable with k categories (employment status and geographic prefecture), k-1 dummy variables (0, 1) were obtained and used as binary in the analysis. Binary variables (gender, existence of chronic health condition, public health insurance, private health insurance and urbanity status of permanent residence) were

treated as such.

Even though the OLS Models was the most widely used method of compositional data analysis [34] in the past, it has been subjected to criticism, mostly because: a) it does not ensure that the fitted values lie in the unit interval and b) the assumptions of normality and heteroscedasticity of the residuals are violated [30]. Violating the homoscedasticity (of the residuals) assumption means that the standard errors and hence the significance tests and confidence intervals will be incorrect [35]. Further, if the residuals are not normally distributed, statistical inferences based on t-statistics or the F-statistic cannot be trusted [36], especially when the sample size is small.

However, with only two categories, as in the case this paper presents, compositional data are (essentially) univariate [37] and can be addressed by means of standard statistical techniques such as OLS [38, 39].

Nevertheless, in order to check the suitability of the OLS Model, I examined whether the fitted values lie in the unit interval. Next, the model was tested as to the normality and homoscedasticity of the standardized residuals using the Jarque-Bera test of normality and Brown & Forsythe test of homoscedasticity respectively. Furthermore, the model was tested as to the existence of specification error through the Ramsey RESET test.

The R (The Compositional Package written by Tsagris and Athineou), Stata and SAS statistical software packages were used for the analysis.

3. Results

Based on the twice minimization of the Kullback-Leibler divergence, the optimal value of the data-based power transformation (α-transformation) of the compositional data is equal to 1. Therefore the data could be treated as they were Euclidean, ignoring the compositional constraint [40, 41].

Additionally, the response variable does not follow the Beta Distribution (Table 3). Therefore, the BR Model was not performed.

Table 3. Goodness of Fit Tests for Beta Distribution

Test	p
Kolmogorov-Smirnov	<0.001
Cramer-von Mises	<0.001
Anderson-Darling	<0.001

According to the Jarque-Bera test, the response variable is normally distributed (p=0.19). The Jarque Bera test was used based on the fact that the literature indicates that its power is higher than the power of Anderson–Darling, Cramer–von Mises, and Kolmogorov-Smirnov tests of normality [42].

Based on the OLS Model the combined utilization of private health services is determined by education, total utilization of healthcare services and urbanity status of permanent residence (Table 4).

Table 4. Results of the OLS Model

Proportion of Both Private and Public Health Services Utillization	Coefficient	p	95% C.I
Education	0.077	<0.001	0.034-0.200
Total Utilization of Healthcare Services	-0.021	0.040	(-0.041)-(-0.001)
Urbanity Status of Permanent Residence	0.093	0.002	0.035-0.152
Constant	0.362	<0.001	0.261-0.463

The fitted values lied in the interval of [0.33, 0.66].

As the Jarque-Bera test of normality returns a p-value of 0.06, the residuals follow the normal distribution.

The Brown & Forsythe homoscedasticity test returned p=0.07, therefore the residuals have constant variance.

According to the Modified Hosmer & Lemeshow Test (p=0.93) the model has a good fit to the data.

Finally, based on the RESET Test (p=0.60) the model does not suffer from specification error.

Based on the aforementioned tests, the OLS Model shows a good fit to data.

However, the QML Model (Table 5) also shows a good fit to data.

Table 5. Results of the QML Model

Proportion of Both Private and Public Health Services Utillization	Coefficient	p	95% C.I
Education	0.315	<0.001	0.138-0.491
Total Utilization of Healthcare Services	-0.086	0.04	(-0.168)-(-0.003)
Urbanity Status of Permanent Residence	0.381	0.002	0.140-0.623
Constant	-0.564	0.01	(-0.98)-(-0.147)

As the Jarque-Bera test of normality returns a p-value of 0.06, the residuals follow the normal distribution.

The Brown & Forsythe homoscedasticity test returned p=0.07, therefore the residuals have constant variance.

Finally, based on the RESET Test (p=0.8) the model does not suffer from specification error.

Based on the aforementioned tests, the QML Model shows a good fit to data.

Based on Table 6, the marginal effects of the OLS Model are comparable to those of the QML Model.

Table 6. Marginal Effects

Variable	OLS Model	QML Model
Education	0.077 (p<0.001)	0.076 (p<0.001)
Total Utilization of Healthcare Services	-0.02 (p=0.040)	-0.02 (p=0.038)
Urbanity Status of Permanent Residence	0.093 (p=0.002)	0.092 (p=0.002)

4. Conclusions

Healthcare-seeking behavior is considered as being multilevel. The first stage of the decision-making process involves the choice of whether to seek medical care or not, while subsequent stages involve the choice of the type of provider and the frequency of utilization [6].

Discrete choice models usually analyze the types of health services consumed, while continuous choice models usually analyze the intensity of healthcare consumption and predict the quantities of health care services consumed [43].

Studying discrete choice in the case of health systems where private and public sectors co-exist, requires the inclusion of a combination of alternatives as an additional mutually exclusive category in the consideration set [44].

However, since several studies focus on a single visit [45, 46] the combination of alternatives is not observed. In addition, in several studies combinations are excluded from the analysis [47], although the use of private and public health services is interdependent [48].

On the other hand, in most studies on continuous choice, the intensity of health service consumption is measured by health expenditure [49]. While health expenditure analysis may serve as proxy of quantity, it focuses mainly on units such as medical visits or operations that have prices attached [50] and thus does not always allow the accurate study of the total quantity of health care services consumed.

Furthermore, analyses of the number of visits to a certain provider that are based on count data models (zero-inflated or hurdle models), depending on the characteristics of the health system (referral system, waiting lists etc.) [51] do not take into account the relative information that must be contained in the choice variable (the response).

The fractional data approach attempts to reduce the problem mentioned above by expressing the choice of a certain provider (measured as a number of visits) relatively to the total health services utilization.

The advantage of this approach is that the derived continuous choice variables (proportions with values in the unit interval) incorporate information not only on the alternatives, but also on the extent to which they are selected. In fact, since the non-use and the exclusive use are included in both the discrete and the continuous choice, these variables are differentiated by the values in the (0, 1) interval i.e. the combined use of alternative sectors' health facilities (in the discrete choice, a single value corresponds to the combined use).

Although the idea of expressing visits in relation to the total utilization is not new [52], this analysis highlights that the combined use of alternative providers is a considerable percentage of total health services utilization that should not be ignored.

Finally, this analysis also points out the need for studying the effect of total utilization on health care provider choice [33] and shows that simple use models, such as OLS, can effectively study patient's selection of healthcare provider [53].

To summarize, the use of alternative healthcare services is a fraction of total health care utilization with values in the unit interval. Values in the interval of (0, 1) express a combined utilization of healthcare services of alternative sectors. The use of the compositional data approach allows for studying the extent of the combined use of healthcare services.

Suggested methods of fractional data analysis include the Dirichlet Models (or Beta Models), the QML Models; the OLS Models consist of an alternative method of fractional data analysis.

Even though the OLS Model was the most widely used method of fractional data analysis in the past, it has been subjected to criticism [54], since its assumptions are violated.

However, *"the normal regression model estimator performs rather well even when the true model is not normal. If this is true then researchers need not worry about the theoretical requirement that the domain of the true density function is the unit simplex, and they should continue to use the normal model and its estimator. One difficulty with this approach is that, if the normal model is incorrect and there exists heteroskedasticity, the wrong formula for the standard errors will be used, whereas the above results are based upon the true standard errors"* [55].

In the case of combined use of private healthcare services, the OLS Model gives satisfactory results. Based on the statistical tests of the OLS Model, this specific model presents a good fit to data, while regression assumptions are not violated.

In addition, the marginal effects of the aforementioned model are similar to those of the QML Model.

The results of the study confirm existing international literature, according to which, in the case of proportions, OLS Models can provide satisfactory results [56, 57].

Even though the dilemma facing researchers to transform or not to transform [58] does not disqualify the use of log-ratio [41], the transformed normal distribution model suffers from an interpretation problem because the expected value of Y is not a simple transformation of the expected value of the transformed response [59].

REFERENCES

[1] Kernick, D. P., 2003, Introduction to health economics for the medical practitioner., Postgrad Med J, 79(929), 147-150.

[2] Steinwachs, D. M. and Hughes, R. G., 2008, Health services research: Scope and significance. In R. G. Hughes (Ed.), Patient safety and quality: An evidence-based handbook for nurses, 1-163-1-177, Rockville (MD): Agency for Healthcare Research and Quality.

[3] Tulchinsky, T. H. and Varavikova, E., 2009, The new public health, San Diego: Elsevier and Academic Press.

[4] Mwabu, G., 2009, The demand for health care. In G. Carrin,

H.K. Buse, K. Heggenhougen, and S.R. Quah, (Eds.), Health systems policy, finance and organization, 245-249, Oxford: Elsevier and Academic Press.

[5] Earl, P. E. and Kemp, S., 1999, The Elgar companion to consumer research and economic psychology, Northampton: Edward Elgar Publishing, Inc.

[6] Jack, W., 1999, Principles of health economics for developing countries. Washington: World Bank.

[7] Gertler, P., Locay, L. and Sanderson, W., 1987, Are user fees regressive? The welfare implications of health care financing proposals in Peru., J Econom, 36(1-2), 67-88.

[8] Cameron, A. C., Trivedi, P. K., Milne, F. and Pigott, J. A, 1988, Microeconomic model of the demand for health care and health insurance in Australia., Rev Econ Stud, 55(1), 85-106.

[9] Rodriguez, M. and Stoyanova, A., 2004, The effect of private insurance access on the choice of GP/Specialist and public/private provider in Spain., Health Econ, 13(7), 689-703.

[10] Dow, W. H., 1999, Flexible discrete choice demand models consistent with utility maximization: An application to health care demand., Am J Agric Econ, 81(3), 680-685.

[11] Yip, W. C., Wang, H. and Liu, Y., 1998, Determinants of patient choice of medical provider, A case study in rural China., Health Policy Plan, 3(3), 311-322.

[12] Hensher, D.A. and Johnson, L.W., 1981, Applied discrete choice modeling. New York: John Wiley & Sons, Inc.

[13] Orturaz, J. dD. and Willumsen, L. G., 2001, Modelling Transport. West Sussex: John Wiley & Sons, Inc.

[14] Penchansky, R. and Thomas, J. W., 1981, The concept of access: definition and relationship to consumer satisfaction., Med Care, 19(2), 127-140.

[15] O'Donnell, O, Van Doorslaer, E., Wagstaff, A. and Lindelow, M., 2007, Analyzing health equity using household survey data: A guide to techniques and their implementation. Washington DC: World Bank.

[16] Jones, A. M., 2006, The Elgar companion to health economics, Northampton: Edward Elgar Publishing, Inc.

[17] Train, K., 2003, Discrete choice methods with simulation. Cambridge: Cambridge University Press.

[18] Cascetta, E., 2009, Transportation systems analysis: Models and applications. New York: Springer.

[19] Hensher, D. A., Rose, J. M. and Greene, W. H., 2005, Applied choice analysis: A primer. Cambridge: Cambridge University Press.

[20] Anderson, D. R., Sweeney, D. J., Williams, T. A, Camm, J. D. and Cochran, J.J., 2015, Essentials of statistics for business and economics. Stamford: Cengage Learning.

[21] Agresti, A., 2002, Categorical data analysis. New Jersey: John Wiley and Sons, Inc.

[22] Folz, D. H., 1996, Survey research for public administration. London: SAGE.

[23] de Gruijter, J., Brus, D., Bierkens M. and Knotters, M., 2006, Sampling of natural resource monitoring. New York: Springer.

[24] Ü stün, T.B., Chatterji, S., Villanueva, M., Bendib, L., Ç elik, C., Sadana, R., Valentine, N., Ortiz, J., Tandon, A., Salomon, J., Cao, Y., Wan Jun, X., Ô zaltin, E., Mathers, C. and Murray, C. J. L., 2001, WHO multi-country survey study on health and responsiveness 2000-2001, GPE Discussion Paper 37 http://www.who.int/healthinfo/survey/whspaper37.pdf.

[25] Mosialos, E., Allin, S. and Davaki, K., 2005, Analysing the Greek health system: A tale of fragmentation and inertia., Health Econ., 14, S151-S168.

[26] Kyriopoulos, J., Lionis, C., Dimoliatis, I., Merkouris, B. P., Oikonomou, C., Tsakos, G. and Philalithis, A. and Working Group: Antonakis, N., Georgoussi, E., Dolgeras, L., Domenikos, C., Zavras, Th., Kontos, D., Kounari, C., Beazoglou, T., Boursanidis, C., Symeonidis, S., Souliotis, K., Tsalikis, G., Vasilikiotis, N., Vrachni, Tz., 2000, Primary health care as a cornerstone for health reform in Greece., Primary Health Care, 12(4), 169-188 (in Greek).

[27] Filzmoser, P., Hron, K.and Reimann, C., 2009, Univariate statistical analysis of environmental (compositional) data: Problems and possibilities., Sci Total Environ, 407(23), 6100-61088.

[28] Aitchison, J., 1986, The Statistical analysis of compositional data. New York: Chapman & Hall.

[29] Smithson, M. and Verkuilen, J., 2006, A better lemon squeezer? Maximum likelihood regression with beta-distributed dependent variables., Psychol Methods, 11(1), 54-71.

[30] Ferrari, S. and Cribari-Neto, F., 2004, Beta regression for modelling rates and proportions., J Appl Stat, 31(7), 799-815.

[31] Cook, A. A. Jr., 1971, A note on estimating proportions by linear regressions. Santa Monica: RAND.

[32] Papke, L. E. and Wooldridge, J. M., 1996, Econometric methods for fractional response variables with application to 401(K) plan participation rates., J Appl Econ, 11(6), 619-632.

[33] Aitchison, J., 1982, The statistical analysis of compositional data., J R Stat Soc Series B Stat Methodol, 44(2), 139-177.

[34] Kieschnick, R. and McCullough, B. D., 2003, Regression analysis of variates observed on (0, 1): Percentages, proportions and fractions., Stat Modelling, 3(3), 193-213.

[35] Cohen, J., Cohen, P., West, S. G. and Aiken, L. S., 2003, Applied multiple regression/correlation analysis for the behavioral sciences. New Jersey: Lawrence Erlbaum Associates.

[36] Crown, W. H., 1998, Statistical models for the social and behavioral sciences: multiple regression and limited dependent variable models. Westport CT: Praeger Publishers.

[37] Billheimer, D., 2002, Compositional data. In A. H. El-Shaarawi, and W.W. Piegorsch (Eds.), Encyclopedia of Environmetrics Vol. 1, 391-399, Chichester: John Wiley and Sons, Inc.

[38] Lipsmeyer, C. S., Philips, A. Q., Rutherford, A. and Whitten, G. D., 2017, Comparing dynamic pies: A strategy for modeling compositional variables in time and space., Political Science Research and Methods, 1-18.

[39] Katz, J.N. and King, G., 1999, A statistical model for multiparty electoral data., Am Polit Sci Rev, 93(1), 15-32.

[40] Tsagris, M., Preston, S. and Wood, A. T. A., 2016, Improved classification for compositional data using the α-transformation., J Classif, 33(2), 243-261.

[41] Tsagris, M., 2015, Regression analysis with compositional data containing zero values., Chil J Stat, 6(2), 47-57.

[42] Steinskog, D. J., Tjøstheim, D. B. and Kvamstø, N. G., 2007, A cautionary note on the use of the Kolmogorov-Smirnov test for normality., Mon. Wea. Rev., 135, 1151-1157.

[43] Adhikari, S. R., 2011, A methodological review of demand analysis: an example of health care services., EJDI, 13 & 14(1-2), 119-130, Combined Issue.

[44] Rani, M. and Bonu, S., 2003, Rural Indian women's care-seeking behavior and choice of provider for gynecological symptoms., Stud Fam Plann, 34(3), 173-185.

[45] Habtom, G. K. and Ruys, P., 2007, The choice of health care provider in Eritrea., Health Policy, 80, 202-217.

[46] Lim, M-K., Yang, H., Zhang, T., Feng, W. and Zhou, Z., 2004, Public perceptions of private health care in socialist China., Health Aff (Millwood), 23(6), 222-234.

[47] Ha, N. T. H., Berman, P. and Larsen, U., 2002, Household utilization and expenditure on private and public health services in Vietnam., Health Policy and Plan, 17(1), 61-70.

[48] Castiglione, D., Lovasi, G. S. and Carvalho, M., S., 2017, Perceptions and uses of public and private health care in a Brazilian Favela., Qual Health Res 00(0), 1-14.

[49] Amaghionyeodiwe, L. A., 2018, Poverty, user charges and health care demand in Nigeria., IJBED, 6(2), 37-50.

[50] Gerdtham, Ulf. G. and Jönsson, B., 1991, Price and quantity in international comparisons of health care expenditure., Appl Econ, 23, 1519-1528.

[51] Winkelmann, R., 2008, Econometric analysis of count data. New York: Springer.

[52] McClellan Buchanan, G., Rubenstein Gardenswartz, C. A. and Seligman, M. E. P., 1999, Physical health following a cognitive-behavioral intervention., Prevention and Treatment, 2, Article 10.

[53] Hess, S. and Daly, A., 2014, Handbook of choice modelling. Northhampton: Edward Elgar.

[54] Paolino, P., 2001, Maximum likelihood estimation of models with beta-distributed dependent variables., Polit Anal, 9(4), 325-346.

[55] Woodland, A. D., 1979, Stochastic specification and the estimation of share equations., J Econom, 10(3), 361-383.

[56] McDonald, J., 2009, Using least squares and tobit in second stage DEA efficiency analyses., Eur J Oper Res, 197(2), 792-8.

[57] Hoff, A., 2007, Second Stage DEA: Comparison of approaches for modelling the DEA score., Eur J Oper Res, 181(1), 425-435.

[58] Ranganathan, Y. and Borges, R. M., 2011, To transform or not to transform: That is the dilemma in the statistical analysis of plant volatiles. Plant Signal Behav, 6(1), 113-116.

[59] Hossain, A, Rigby, R., Stasinopoulos, M. and Enea, M., 2016, Centile estimation for a proportion response variable., Statist. Med., 35, 895-904.