University of Economics in Katowice

Journal of Economics and Management

ISSN 1732-1948 Vol. 35 (1) • 2019

**Saheed O. Olayiwola**

https://orcid.org/0000-0003-2500-0229

Department of Economics
Federal University of Technology, Akure
saheedolayiwola@yahoo.com

**Bayo L.O. Kazeem**

Department of Business Administration
Elizade University
Ilara-Mokin, Ondo State
oyerokazeem@yahoo.com

# Count data modelling of health insurance and health care utilisation in Nigeria

doi: 10.22367/jem.2019.35.06

Accepted by Editor Ewa Ziemba | Received: March 7, 2018 | Revised: August 27, 2018; October 25, 2018; November 20, 2018 | Accepted: December 5, 2018.

## Abstract

**Aim/purpose** – Estimation of the model of interdependent demand for health insurance and health care utilisation involves issues of stochastic dependence between health insurance and health care utilisation. This study explored a count data estimation technique to determine the most appropriate estimation method for the interdependence of health insurance and health care demand in Nigeria.
**Design/methodology/approach** – The study employed Hidayat and Pokhrel (2010) framework to choose among the six alternatives of two classes of count data model. The data for the study were collected using a purposive sampling survey in the six geopolitical zones in Nigeria.
**Findings** – The results showed that the general method of moments (GMM) estimator is preferable to model the determinants of medical care consumption with health insurance. Price of health care services is positively related to medical care consumption with health insurance and social health insurance. The income-medical care relationship indicated that medical care services are inferior good under private health insurance and a normal good with social health insurance during sick period.
**Research implications/limitations** – The implication of this study is that the estimation method that accommodates endogenous regressors is the appropriate estimation technique for the interdependence of health insurance and health care utilisation. The limitation of this study is that the recall period was just six months prior to the survey.
**Originality/value/contribution** – The study revealed that the estimation techniques for the interdependence of health insurance and health care utilisation must recognised the influence of individual and household characteristics on the decision to purchase health

insurance and health care consumption. Hence, diagnostics tests are require to choose the most appropriate estimation technique.

**Keywords:** health insurance, health care utilisation, count data model.
**JEL Classification:** I130.

## 1. Introduction

Three attributes complicate the estimation of the interdependent demands for health insurance and health care. First, insurance is not distributed randomly and possibly endogenous to the health care decision. This may lead to potential biases in the estimation of health care demand if left uncontrolled. Second, the differences in health care use across insurance regimes cannot be addressed with a single parameter. Insurance likely modifies the relationship between the socio--economic variables and health care use by providing access to an entirely different system of care. Third, the use of health care is discrete and non-negative in the form of a count of services over a period of time (Koç, 2005). Therefore, the endogeneity of health insurance complicates the estimation of the relationship between insurance and health care use. These complexities are due to the underlying behaviours driving health care utilisation that may have implications for the choice of the model (Vera-Hernández, 1999; Waters, 1999). Though, choosing a model appropriate for estimating health care demand is a difficult process which was poorly documented in the health economics literature (Hidayat & Pokhrel 2010). Nevertheless, when a dependent variable have only non--negative integer values, count data models provides appropriate estimation techniques (Jones, 2000; Vera-Hernández, 1999).

Econometric implementation of the model of interdependent demand for health insurance and health care may involves problems of discreteness of choice, selectivity and stochastic dependence between health insurance and utilization (Cameron, Trivedi, Milne, & Piggott, 1988). Discrete choice was due to government regulation of insurance companies that provide over 90% of the health insurance. Therefore, the direction of the bias due to unobserved heterogeneity is unclear a priori. However, several studies control for observed health status and other individual characteristics in a multivariate regression framework, instrumental variables techniques or panel data to account for the endogeneity of insurance coverage. This modelling problem was similar to other choice models involving jointness of optimal discrete and continuous choices (Koç, 2005). Examples are Cameron et al. (1988) who modelled interdependent demand for health insurance and health care under uncertainty and Hidayat & Pokhrel (2010) who examined the selection of an appropriate count data model for modelling health insurance and health care demand in Indonesia. These modelling approaches have their strengths and weaknesses.

Therefore, within the context of Hidayat & Pokhrel study (2010), this paper explores count data modelling technique with econometric specification tests to determine the appropriate estimation technique for the interdependence of health insurance and health care demand in Nigeria.

The remainder of the study is structured as follows: section 2 presents theoretical backgrounds while section 3 contains research methodology. Section 4 discusses the empirical results and section 5 offers concluding remarks.

## 2. Theoretical backgrounds

### 2.1. Literature review

Liu, Nestic, & Vukina (2012) employed invoices for hospital services from a regional hospital in Croatia to test for adverse selection and moral hazard with three categories of patients: with no supplemental insurance, who bought it and who are entitled to it for free. The identification procedure relies on the premise that the difference in the observed medical care consumption between the patients who bought the insurance and those entitled to free insurance is caused by pure selection effect; whereas the difference in health care consumption between the group that received free insurance and the group that has no insurance is due to moral hazard. The estimation was done with the use of matching estimators which compares the outcomes of participants with those of matched non-participants where matches are chosen based on the similarity in observed characteristics. The framework assumes two potential outcomes that represent control (no treatment and treatment states) and two groups of individuals. Since this study was a control experiment, the results may depend on the individual's state of health during the experiment period and hence create an endogeneity problem.

Bajari, Hong, Khwaja, & Marsh (2006), estimate a structural model of the demand for health insurance and medical care with a two-step semi-parametric method to separate adverse selection from moral hazard in health care. A standard functional utility function was specified with data from health and Retirement Study. The study specified a model of endogenous consumer demand for health insurance and medical utilisation. This allows for heterogeneity in the distribution of latent health status using semi-parametric estimator to recover the parameters of the utility function and non-parametric test for adverse selection. The authors also used a distribution-free test, to test for adverse selection using Kolmogorov–Smirnov test statistics. The major advantage of this strategy is that estimations do not rely on parametric assumptions about the latent health distribution in estimating the parameters of the model. Since this study was a static empirical framework, individual preferences and health risk can affect insurance decisions and future health investments. Hence, the unobserved nature of individual preferences and health risk can cause heterogeneity problems.

Bolhaar, Lindeboom, & Klaauw (2008) specified and estimated a dynamic panel data model using the Living in Ireland Survey to deal with the problem involved in a cross-section data in a dynamic analysis of the demand for health insurance and health care. The models allowed for individual specific effects that captured heterogeneity in preferences and health risk using pooled OLS as a baseline case. The pooled OLS estimates ignore state dependence, but include time-invariant regressors, such as level of education. The study also estimates a static fixed effects model that allows for unobserved household (insurance decision) and individual (health care utilisation) specific effects, but ignores the dynamic structure of the process. General Methods of Moment (GMM) was used to estimate dynamic panel data models that include unobserved effects and state dependence. Different estimation methods were used because studies on health insurance and medical care utilisation were based on cross-sectional analyses and different estimation methods can determine whether the results conform to previous findings and how the deviations from the models change the results.

The problem of unobserved heterogeneity was confronted by the RAND Health Insurance Experiment (HIE) in the United States. The HIE was a social insurance randomised experiment for estimating the elasticity of demand for health care (Newhouse, Sloss, Manning, & Keeler, 1993; Zweifel & Manning, 2000). The concern was that estimates based on observational studies are often systematically biased in their estimates of the responses to insurance coverage (Cutler & Zeckhauser, 2000). About 6,000 people were randomised to different insurance plans in six areas over a three to five-year period in the early 1970s. The insurance plans varied in contractual levels of cost sharing. The elasticity estimates were computed by comparing utilisation in different plans. The implications of this study for health policy may be limited today due to the type of insurance studied (mostly indemnity) and the timing of HIE.

Cameron et al. (1988) addressed the problem of endogeneity and unobserved heterogeneity in a joint model of demand for health care and health insurance, using 1977-1978 Australian Health Survey. A Negative binomial model (a generalisation of the Poisson) and zero-inflated binomial model in a count data modelling was employ to model utilisation conditional on insurance choice. The use of count data modelling was justified on the basis that in estimating the effects of health insurance on the demand for health care, it is important to establish whether the demand variable is generated as a discrete and mutually exclusive choice (e.g. types of provider visited in the event of illness) or is in the form of count or rate (e.g. number of visits made to particular provider). Koç (2005) also examined the effect of insurance on the demand for health care among consumers of similar health (called health-specific moral hazard effect). Using the 2000 Medical Expenditure Panel Survey, he analysed the variation in the moral hazard effect across health subpopulations in the demand for inpatient and outpatient services. An endogenous switching model (also known as type-5 Tobit) for count data was used to deal with the endogeneity of insurance, the change of insurance regime and the discreteness and the non-negativity of the use of health care. Cameron et al. (1988) and Koç (2005) were an improvement in solving the problem of endogeneity of insurance with health care utilisation using different estimation method of count data modelling. However, failure to evaluate the overall specification of the model left problems like heteroscedasticity and over-dispersion unaddressed.

Hidayat & Pokhrel (2010), on the selection of appropriate count data model for modelling health insurance and health care demand in Indonesia, conducted thorough econometric specification tests to address all possible problems that may occur from endogeneity of insurance with health care utilisation. They compared estimation from Hurdle Negative Binomial (HNB) and GMM and concluded that HNB estimation performs better than GMM estimation. The thorough econometrics specification tests carried out in the study significantly reduced the estimation problems that endogeneity of insurance, individuals, and households' specific characteristics may create.

## 2.2. Theoretical framework

The study adopted contract theory within the context of health insurance in a developing economy. It is assumed that individuals seeking to enter into health insurance contract are not selected at random and individual characteristics, such as health status may influence the decision to enter into a contract and thus create a self-selection bias. In other words, the individual may have information about the probability of poor health, which the insurer cannot observe. Thus, individuals with low expectations about their future health status may have an incentive to select insurance coverage. It is further assumed that under uncertainty, risk-averse individuals demand risk-bearing goods, such as health insurance, to safeguard their income against possible shocks. Health is assumed to be a choice variable because it is a source of utility. Individuals value health, with health care as a means of producing health. Therefore, individuals first choose their insurance and then choose their health care utilisation when ill. The related uncertainty under this scenario is with respect to future health status at the time the insurance policy is chosen. Also, insurance is purchased because the expected value of the additional health care and other consumer commodities if ill exceeds the expected cost of paying the insurance premium if healthy. Thus, the demand for health care services, unconditional of insurance choice can be written as a non-linear equation of the form:

$$E\left[m_i(s)\right] = \exp(\eta P_m + \varepsilon Y^{'} + \sum_{j=1}^{J} \Phi_{ji} D_j + \alpha_i Z_i + \mu_2) \tag{1}$$

Equation (1) is a demand equation for medical care consumption given health insurance status. The dependent variable $[m_i(s)]$ in the equation is the number of times of consuming health care such as physician consultations, drug use, number of inpatient and outpatient days etc. This variable is in the form of count (i.e. 1, 2, 3, … , ∞). This motivates the use of count data estimation model. ($Z_i$) is a vector of individual household characteristics that may be important in health care decision, like household size, education, marital status and employment status. ($P_m$) is the price of health care measured by the co-insurance rate multiplied by individual monthly total health expenditure, ($Y'$) is the income in time two (assume illness occurred in time two) and ($D_j$) are dummy variables for the ($j_{th}$) insurance form (the value of j is between 0 and 1 where 0 indicates non-insured and 1 indicates insured). Individuals were also categorised under social health insurance and private health insurance. The implication of this model is that from the demand equation if the price of the health care services is lower more of it will be demanded. Thus, we would expect that ($\Phi_{ji}$) is larger for the insurance policy ($j$) that is more generous. This is the moral hazard effect of health insurance and ($\mu_2$) is the error term that represents other unobserved characteristics.

## 3. Research methods and procedures

### 3.1. Estimation technique

Equation (1) measures the determinants of demand for health care services by the insured. Since equation (1) is an exponential equation, coefficients from this equation were interpreted as elasticity. Hence, if the coefficient ($P_m$) is less than one (i.e. inelastic) it indicates the existence of moral hazard. Three versions of equation (1) were estimated. These are whether an individual has health insurance or not, for social and private health insurance. Specification tests were employed to choose among the two classes of count data models. The first class is characterised by a primary equation with a discrete dependent variable. This includes standard count data models such as restricted Poisson, negative binomial, zero-inflated negative binomial and hurdle models. The second class extends the features of the first class to accommodate endogenous regressors. This includes Instrumental Variables (IV) and Generalised Method of Moments (GMM) techniques. The models were used in the estimation of equation (1) with maximum likelihood techniques choosing robust standard error procedures in anticipation of the misspecification of the true (but unknown) population density.

Three main steps involve in choosing the most appropriate econometric technique among the six alternatives of two classes of count data model. The first class is the standard count data models such as restricted Poisson, negative binomial, zero-inflated negative binomial and hurdle models. The second class is the Instrumental Variables (IV) and Generalised Method of Moments (GMM) techniques which extends the features of the first class to accommodate endogenous regressors. The first specification test checks for endogeneity using Hausman specification tests (Wu–Hausman and Durbin–Wu–Hausman). A significant difference between coefficients from ML and GMM or IV, suggests that the null hypothesis of exogeneity should be rejected. This implies that either IV or GMM estimator is necessary. Pagan and Hall's test for heteroscedasticity (Pagan & Hall, 1983) was used to determine the choice between IV and GMM estimators. The rejection of the null hypothesis of homoscedasticity suggests that GMM is preferable to IV. The consistency of the endogeneity test, as well as coefficient estimates of IV and GMM, depend on the validity of the instruments. The instruments are the variables that have an impact theoretically and conceptually, on the suspected endogenous variable but that do not affect the dependent variable. The variables include individual households and socio-economic characteristics. Identification of the effect of insurance status on health care demand is achieved if the Z's are uncorrelated with the structural error but correlated with the endogenous regressors, that is, health insurance variable.

To evaluate whether potential instruments are weak and the instruments are orthogonal to the error process, several tests were employed. First, the relevance of the instruments (to suspected endogenous variables) was assessed by evaluating the $R^2$ value and the $F$-test for the joint significance of the instruments in the first-stage regressions. The first-stage regressions are reduced-form regressions of the endogenous variables on the full set of instruments and other exogenous regressors. If the models have more than one suspected endogenous variable, relying only on $R^2$ and $F$- statistics may not be enough to detect the relevance of the instruments. Therefore, a Shea partial $R^2$ measure that takes correlations among the instruments into account will be appropriate (Shea, 1997; Staiger & Stock, 1997). The smaller the value of the partial $R^2$, the more inconsistent the IV estimates will be whenever the instruments are not perfectly exogenous. Even when the instruments are exogenous, a small value of the partial $R^2$ will mean increased asymptotic standard errors and reduction in the power of the $F$-test.

Second, the validity of the instruments was tested by an over-identification test (Windmeijer & Santos-Silva, 1997). Hansen's J-statistics and the Sargan statistics for GMM and IV were employ respectively (Baum, Schaffer, & Stillman, 2003). The joint null hypothesis of Hansen and Sargan tests is that the excluded instruments are valid instruments and their correctly excluded from the estimated equation. Finally, to satisfy an orthogonal requirement of the instruments (i.e. that the instruments are exogenous), a subset of the instruments are tested using the $C$-statistics (Baum et al., 2003), that allow a subset of the original set for exogeneity conditions to be tested. Count data models that ignore endogeneity are employ when the null hypotheses of exogenous regressors are accepted.

## 3.2. Data

The data for the study were collected using a purposive sampling survey in the six geopolitical zones in Nigeria. The six geopolitical zones are South-West, South-East, South-South, North-West, North-East, and North-Central. One state with a large presence of formal sector workers was chosen from each zone. This choice was based on the fact that the former sector workers are mostly covered by health insurance in Nigeria. Lagos State was chosen in the South-West, Imo in the South-East, Rivers in the South-South, Kaduna in the North-West, Adamawa in the North-East and Abuja in the North-Central. The survey was conducted in hospitals, government parastatals, private companies, and households. The target population was formal sector employees (private or public) and informal sector workers with or without health insurance coverage. The tool for the study is a self-designed 48 items questionnaire with questions on households' socio-demographic characteristics, health insurance status, health status, health care expenditures and health care utilisation.

### 3.3. Description of variables

The dependent variable in equation (1) indicates the number of times of consuming health care services by individuals with health insurance and otherwise. The independent variables include the price of health care services, income during illness defined as individual's monthly income plus the proportion of health expenditure paid by insurance, health insurance status and the type of health insurance held by individuals. $Z_i$ is also individual's household characteristics that can influence the demand for health care services such as household size, level of education, employment status and marital status. Appendix I shows the variables and their definitions.

## 4. Research findings and discussion

### 4.1. Descriptive statistics and demographics

Table 1 shows the descriptive statistics of the variables employed in the study.

**Table 1.** Summary statistics of the variables used for estimation

| Variables | | Obs | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|---|
| *1* | | *2* | *3* | *4* | *5* | *6* |
| Married1 | Single | 1,051 | 0.4757 | 0.4997 | 0 | 1 |
| Married2 | Married | 1,051 | 0.4738 | 0.4996 | 0 | 1 |
| Married3 | Divorce/Separated | 1,051 | 0.0105 | 0.1018 | 0 | 1 |
| Married4 | Widowed | 1,051 | 0.0399 | 0.1960 | 0 | 1 |
| Male1 | Male = 1 | 1,051 | 0.5119 | 0.5001 | 0 | 1 |
| Male2 | Female = 1 | 1,051 | 0.4881 | 0.5001 | 0 | 1 |
| Age | | 1,051 | 32.6870 | 11.3344 | 16 | 80 |
| FMTYPE1 | Monogamy = 1 | 1,051 | 0.7431 | 0.4371 | 0 | 1 |
| FMTYPE2 | Polygamy = 1 | 1,051 | 0.2569 | 0.4371 | 0 | 1 |
| FMHEAD1 | Father = 1 | 1,051 | 0.9125 | 0.2828 | 0 | 1 |

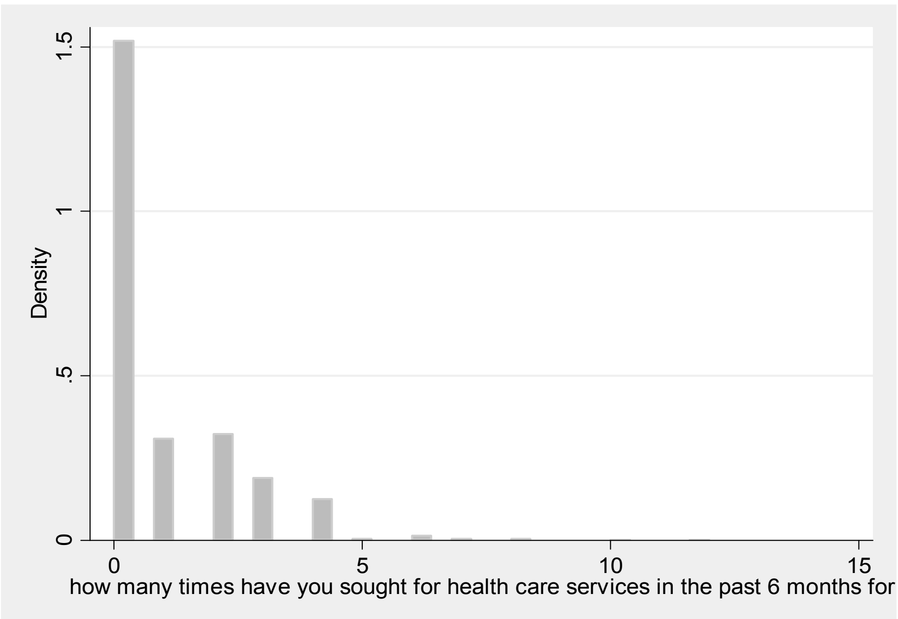
**Table 1 cont.**

| *1* | *2* | *3* | *4* | *5* | *6* |
|---|---|---|---|---|---|
| FMHEAD2 Mother = 1 | 1,051 | 0.0875 | 0.2828 | 0 | 1 |
| FMHEDUC1 No Formal Schl. = 1 (Father) | 1,051 | 0.0504 | 0.2189 | 0 | 1 |
| FMHEDUC2 Primary Edu = 1 | 1,051 | 0.0428 | 0.2025 | 0 | 1 |
| FMHEDUC3 Sec. Edu = 1 | 1,051 | 0.1570 | 0.3640 | 0 | 1 |
| FMHEDUC4 Post Sec. Edu = 1 | 1,051 | 0.7498 | 0.4333 | 0 | 1 |
| SFMHEDUC1No Formal Schl.= 1 (Mother) | 1,051 | 0.0676 | 0.2511 | 0 | 1 |
| SFMHEDUC2Primary Edu = 1 | 1,051 | 0.0666 | 0.2495 | 0 | 1 |
| SFMHEDUC3Sec. Edu = 1 | 1,051 | 0.1665 | 0.3727 | 0 | 1 |
| SFMHEDUC4Post Sec. Edu = 1 | 1,051 | 0.6993 | 0.4588 | 0 | 1 |
| FMHOCC1 Govt. Worker = 1 | 1,051 | 0.5404 | 0.4986 | 0 | 1 |
| FMHOCC2Form. Pvt Sec Worker = 1 | 1,051 | 0.1408 | 0.3480 | 0 | 1 |
| FMHOCC3 Trader = 1 | 1,051 | 0.0733 | 0.2607 | 0 | 1 |
| FMHOCC4 Transporter = 1 | 1,051 | 0.0447 | 0.2068 | 0 | 1 |
| FMHOCC5 Farmer = 1 | 1,051 | 0.0542 | 0.2266 | 0 | 1 |
| FMHOCC6 Self-Employed = 1 | 1,051 | 0.1094 | 0.3123 | 0 | 1 |
| FMHOCC7 Housewife = 1 | 1,051 | 0.0143 | 0.1187 | 0 | 1 |
| FMHOCC8 Unemployed = 1 | 1,051 | 0.0076 | 0.0870 | 0 | 1 |
| FMHOCC9 Others = 1 | 1,051 | 0.0152 | 0.1224 | 0 | 1 |
| SFMHOCC1 Govt. Worker = 1 | 1,051 | 0.4234 | 0.4943 | 0 | 1 |
| SFMHOCC2 Fom Pvt Sec Worker = 1 | 1,051 | 0.1532 | 0.3604 | 0 | 1 |
| SFMHOCC3 Trader = 1 | 1,051 | 0.1941 | 0.3957 | 0 | 1 |
| SFMHOCC4 Transporter = 1 | 1,051 | 0.0238 | 0.1525 | 0 | 1 |
| SFMHOCC5 Farmer = 1 | 1,051 | 0.0504 | 0.2189 | 0 | 1 |
| SFMHOCC6 Self-Employed = 1 | 1,051 | 0.1075 | 0.3099 | 0 | 1 |
| MEXPFD | 1,051 | 18,415.17 | 12204.4 | 100 | 100,000 |
| MEXPTC | 1,051 | 9,626.948 | 7,214.841 | 200 | 100,000 |
| MEXPHLT | 1,051 | 7,173.292 | 6,497.079 | 50 | 100,000 |
| MEXPORS | 1051 | 9,026.081 | 7,569.926 | 100 | 120,000 |
| MTOTAEXP | 1,051 | 34,784.7 | 25,324.09 | 1500 | 400,000 |
| HINSTATUS1 Non-Insured = 1 | 1,051 | 0.3853 | 0.4869 | 0.000 | 1.0000 |
| HINSTATUS2 Insured = 1 | 1,051 | 0.6147 | 0.4869 | 0.000 | 1.0000 |
| HINSTYPE1 NHIS = 1 | 646 | 0.9087 | 0.2883 | 0.000 | 1.0000 |
| HINSTYPE2 PRCHI = 1 | 646 | 0.0619 | 0.2412 | 0.000 | 1.0000 |
| HINSTYPE3 PERHI = 1 | 646 | 0.0294 | 0.1691 | 0.000 | 1.0000 |
| GHSTATUS | 1,051 | 1.0313 | 1.5832 | 0.000 | 8.0000 |
| COINS | 1,051 | 0.1051 | 0.0185 | 0.1 | 0.5 |
| PRICEHC | 1,051 | 750.6553 | 690.2583 | 5.2540 | 10,508.2 |
| MEDICONSUMP | 1,051 | 0.9058 | 1.4195 | 0.000 | 12 |
| COSTRANS | 951 | 162.5447 | 99.3507 | 20 | 700 |
| MTINCO | 1,051 | 68859.98 | 10,6055.3 | 1000 | 3,000,000 |

Table 1 shows that 61.5% of the respondents have health insurance while 38.5% are without health insurance. On the type of health insurance, NHIS represents compulsory social health insurance for public and formal private sector workers; PRCHI represents private company health insurance. The last insurance category is personal health insurance (PERHI). From the summary, about 90.9% of the respondents use NHIS, 6.2% has PRCHI while about 2.9% are covered by PERCHI. The total monthly income of the respondents' ranges from \$6.25 to \$18,750; with average monthly income being \$430.4. The average price of health care is \$4.7 and average general health status score is about 1.03. Other socio-demographic characteristics shows that about 47.6% are single, about 47.4% are married, about 1.04% are divorced or separated and about 3.9% are widowed. Also, about 80.2% of the respondents have post-secondary education; about 14.4% have secondary school education, about 3.3% have primary school certificate while about 2.9% did not attend any formal school. On respondents' occupation, about 41.1% are government workers, 35.4% are formal private sector workers and about 16% are self-employed, about 2% are housewives and about 0.9% are unemployed. This shows that about 76.5% of the respondents are formal sector workers.

The number of times of medical care consumption enters equation (1) as dependent variable to estimate demand for medical care services. From Table 1, this ranges from 0 to 12. The mean value is about 1 while the variance is about 2. The ratio of the variance and the mean is 2.24. This average indicates that the observed data is over-dispersed. Figure 1 further shows evidence of excess zero of the medical care consumption. It shows the density of zero to be 1.5. This motivates count data estimation technique.

**Figure 1.** Number of times of consuming health care services in the past six months



## 4.2. Model selection

The variable that captures the demand for health care is the number of medical care consumption six months prior to the household survey. The discreteness and non-negativity of this variable require count data modelling. General health status and health insurance status are likely to be endogenous to the demand for medical care services; therefore, we have two possible endogenous variables (health insurance status and general health status). Endogeneity tests were first used to choose between the first and second class of count data models. The endogeneity tests on instrumental variable (IV) estimation of medical care consumption with health insurance, social health insurance, and private health insurance were significant at 1% level (Table 2). This favour the use of count data models that accommodate endogenous regressors, i.e. IV or GMM.

**Table 2.** Endogeneity test

| E Endogeneity test | Medical Care Consumption | | | | | |
|---|---|---|---|---|---|---|
| | Health insurance | | Social health insurance | | Private health insurance | |
| | Statistics | p-value | Statistics | p-value | Statistics | p-value |
| Wu–Hausman | F (2,1015) = 7.5202 | 0.001 | F (2,1014) = 7.082 | 0.001 | F (2,1014) = 6.841 | 0.001 |
| Durbin–Wu–Hausman | $\chi^2((2))$ = 15.346 | 0.001 | $\chi^2((2))$ = 14.465 | 0.001 | $\chi^2((2))$ = 13.979 | 0.001 |

Specification tests shown in Tables 3 and 4 were used to choose between Instrumental Variable and GMM techniques. Pagan and Hall heteroscedasticity tests in Table 3 were used to choose either Instrumental Variable (IV) or GMM estimator for reliability. The Pagan and Hall's test in IV 2SLS and GMM estimates with medical care consumption were $\chi2(2) = 10.975$ with *p-value* = 0.004 and ($\chi2(2) = 10.916$ with *p-value* = 0.004 for health insurance; $\chi2\ (2) = 15.406$ with *p-value* = 0.001 and $\chi2(2) = 12.404$ with *p-value* = 0.002 for social health insurance and $\chi2(2) = 19.765$ with *p-value* = 0.000 and $\chi2(2) = 16.669$ with *p-value* = 0.000 for private health insurance respectively. These show the presence of heteroscedasticity in the estimates. This suggests that GMM estimator is preferable to model the determinants of medical care consumption with health insurance, social and private health insurance.

**Table 3.** Pagan–Hall test of heteroscedasticity

| Tests | Medical Care Consumption | | | | | |
|---|---|---|---|---|---|---|
| | Health insurance | | Social health insurance | | Private health insurance | |
| | Statistics | p-value | Statistics | p-value | Statistics | p-value |
| IV 2SLS | $\chi^2((2))$ = 10.975 | 0.004 | $\chi^2((2))$ = 15.406 | 0.001 | $\chi^2((2))$ = 19.765 | 0.000 |
| GMM | $\chi^2((2))$ = 10.916 | 0.004 | $\chi^2((2))$ = 12.404 | 0.002 | $\chi^2((2))$ = 16.669 | 0.000 |

Tests of $R^2$, partial $R^2$, Shea Partial $R^2$ and Wald-test (of instruments and excluded instruments) of the first stage regression on GMM estimates were employed to test the relevance, validity and orthogonality requirements of the instruments. The $R^2$ shows that the models explained a good proportion of the variation in medical care consumption. The values of Partial $R^2$ and Shea-Partial $R^2$ indicate that the models are well identified. The relevance of the instruments was investigated using $F$-test to determine whether the instruments were correlated with the potentially endogenous variable. The null hypotheses of $F$-tests that the parameters of the covariates were jointly equal to zero were rejected. Hence, the instruments are jointly significant with GMM estimator.

**Table 4.** Tests for the relevance of instruments

| Test statistics | Medical Care Consumption | | | | | |
|---|---|---|---|---|---|---|
| | Health insurance | | Social health insurance | | Private health insurance | |
| | General health status | Health insurance status | General health status | Health insurance status | General health status | Health insurance status |
| Unadjusted $R^2$ | 0.3924* | 0.2086* | 0.3924* | 0.2445* | 0.3924* | 0.1133* |
| Adjusted $R^2$ | 0.3702* | 0.1797* | 0.3702* | 0.2146* | 0.3702* | 0.0809* |
| Partial $R^2$ | 0.3102* | 0.0269* | 0.3102* | 0.0168* | 0.3102* | 0.0431* |
| Shea Partial $R^2$ | 0.0762* | 0.0266* | 0.1513* | 0.0082* | 0.2680* | 0.0372* |
| F-tests: | 17.68* | 7.21* | 17.68* | 8.17* | 17.68* | 3.50* |
| Wald test[a] Wald test[b] | 75.92 | 4.66 | 75.92 | 2.89 | 75.92 | 7.60 |

[a] F (37, 1013).
[b] F-test excluded instruments F (6, 1013).
* Significant at 1%.

The validity of the instruments was performed using Hansen's J-statistics for the over-identifying restrictions and C-statistics for the orthogonality condition. The null hypothesis of correct specification in demand for health insurance, social and private health insurance cannot be rejected. The values of the Hansen's $J$-statistic (GMM-estimates) in medical care consumption were 7.101 ($p$-value = 0.13067), 8.834 ($p$-value = 0.65390) and 8.210 ($p$-value = 0.84106), respectively. The value of C-statistics for the orthogonality condition of the instruments were 3.077 (*p-value* = 0.215), 5.078 (*p-value* = 0.152) and 4.402 (*p-value* = 0.111) which indicate that all instruments are exogenous. The specification tests suggest that the selected instruments (head of the family having post-secondary education, spouse of the family head having post-secondary education, the family head being a government employee, spouse of the family head as a government employee, having inherited the disease, having a chronic disease) were appropriate. The implication of this modeling procedure is that all estimation complexities that can give biase results are significantly minimised. Therefore, the ensuing estimation method, given the available data, is the most appropriate estimation technique.

## 4.3. The demand for medical care by the insured

Table 5 shows the results of demand for medical care services with health insurance, social and private health insurance using GMM estimation technique. General health status is inversely related to medical care consumption under health insurance, social and private health insurance. This means that individuals with bad health status may consume more medical care when covered by health insurance, social or private health insurance. Given individuals health insurance status, the results show that those who are covered by health insurance, social or private health insurance may likely demand for less medical care services. Price of health care services is positively related to medical care consumption with health insurance, social health insurance and a positive effect on private health insurance. This shows that the demand for medical care services by the insured is inelastic. The income (during the sick period) medical care coefficient shows that medical care is an inferior good under health and private health insurance and a normal good with social health insurance during sick period.

**Table 5.** GMM estimation of the determinants of the demand for medical care consumption

| MEDICONSUMP | Health insurance | | Social health insurance | | Private health insurance | |
|---|---|---|---|---|---|---|
| | Coeff[a] | (se)[b] | Coeff[a] | (se)[b] | Coeff[a] | (se)[b] |
| *1* | *2* | *3* | *4* | *5* | *6* | *7* |
| GHSTATUS | −0.0761 | 0.0970 | −0.1354** | 0.0675 | −0.1501** | 0.0518 |
| HINSTYPE | −1.6638 | 1.2914 | −1.3014 | 1.2103 | −2.5929 | 2.0121 |
| InPRICEHC | 0.1592 | 0.1018 | 0.1392 | 0.0968 | 0.0001 | 0.0728 |
| InSICKINC | −0.0918 | 0.1633 | 0.0625 | 0.1278 | −0.0926 | 0.1496 |
| COSTRANS | 0.0006 | 0.0006 | 0.0005 | 0.0006 | 0.0009*** | 0.0005 |

**Table 5 cont.**

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| Single[R] | | | | | | |
| Married2 | 0.0373 | 0.1696 | 0.0142 | 0.1653 | −0.0902 | 0.1095 |
| Married3 | 0.2728 | 0.7378 | 0.0425 | 0.6267 | −0.4916 | 0.4282 |
| Married4 | −0.3349 | 0.3946 | −0.3894 | 0.4086 | −0.7621* | 0.1669 |
| Male[R] | | | | | | |
| Male2 | 0.1881 | 0.1366 | 0.1977 | 0.1514 | 0.0155 | 0.0974 |
| Age | 0.0108** | 0.0054 | 0.0118** | 0.0051 | 0.0103** | 0.0048 |
| Monogamy[R] | | | | | | |
| FMTYPE2 | −0.0789 | 0.1461 | −0.0433 | 0.1321 | 0.0569 | 0.1037 |
| Father[R] | | | | | | |
| FMHEAD2 | −0.2988 | 0.2594 | −0.2587 | 0.2611 | −0.0132 | 0.1645 |
| PostSecondary[R] | | | | | | |
| FMHEDUC1 | 0.0329 | 0.2364 | 0.1436 | 0.2106 | 0.0846 | 0.2033 |
| FMHEDUC3 | −0.1569 | 0.1773 | −0.1154 | 0.1934 | −0.2295 | 0.1494 |
| PostSecondary[R] | | | | | | |
| SFMHEDUC3 | 0.0165 | 0.2282 | 0.0253 | 0.2341 | 0.2143 | 0.1549 |
| Govt-Worker[R] | | | | | | |
| FMHOCC2 | −0.0796 | 0.1425 | −0.0952 | 0.1354 | −0.0673 | 0.1293 |
| FMHOCC3 | −0.0723 | 0.2991 | −0.1266 | 0.3324 | 0.0684 | 0.2336 |
| FMHOCC4 | 0.1241 | 0.2841 | −0.0122 | 0.2281 | −0.1314 | 0.1929 |
| FMHOCC5 | 0.2723 | 0.2401 | 0.3259 | 0.2392 | 0.1666 | 0.2353 |
| FMHOCC6 | 0.3459*** | 0.1840 | 0.2579 | 0.1667 | 0.3615** | 0.1615 |
| Govt-Worker[R] | | | | | | |
| SFMHOCC2 | −0.4595* | 0.1451 | −0.4155* | 0.1421 | −0.5417* | 0.1438 |
| SFMHOCC3 | −0.3722** | 0.1547 | −0.3529** | 0.1562 | −0.5183* | 0.1537 |
| SFMHOCC4 | −0.4019 | 0.3755 | −0.5396 | 0.3290 | −0.8116* | 0.2855 |
| SFMHOCC5 | −0.471*** | 0.2504 | −0.4765** | 0.2378 | −0.3628*** | 0.1953 |
| SFMHOCC6 | −0.4859* | 0.1627 | −0.4920* | 0.1486 | −0.5844* | 0.1432 |
| CONSTANT | 2.16803 | 1.9869 | 0.1855 | 1.4239 | 2.1286 | 1.9311 |
| | $R^2 = 0.1614$ | | $R^2 = 0.2306$ | | $R^2 = 0.0586$ | |
| No of Observations | 1051 | | 1051 | | 1051 | |

*, **, and *** Significant at 1%, 5%, and 10% level, respectively.
[a] Estimated parameters.
[b] Robust standard errors.
[R] Reference group.

The households' socio-demographic characteristics results show that married, divorced or separated and widow consumed more medical care under health insurance, social health insurance and less under private health insurance. Households with the head of household without formal schooling consumed more medical care given health insurance, social and private health insurance while households with a head having a secondary school education consumed less medical care. Also, a household headed by a formal private sector worker consumed more medical care compared to other types of employment.

# 5. Conclusions

## 5.1. Research contribution

This study examined the most appropriate estimation technique for estimating the determinants of the demand for medical care by individuals with health insurance and different types of health insurance using the methodology of Hidayat & Pokhrel (2010). The endogeneity tests results favoured the use of the second class of the count data model that accommodates endogenous regressors. The specification tests carried out to choose between instrumental variable (IV) and general methods of moments (GMM) estimators favoured the use of GMM for the estimation of the demand for medical care consumption by individuals cover by health insurance. The modelling results further confirm the need to take cognisance of households' individual's characteristics in the analysis of health care market. Like other studies in the health care demand analysis, the results favoured estimation technique that accommodates endogenous regressors. The analysis of the demand for health care services revealed that the demand for health care given health insurance and different types of health insurance was inelastic while the income elasticities of demand for health care services show that increase in income during sick period encourages increase utilisation of health care services and other goods and services.

## 5.2. Research implication

The modelling procedure implies that estimation complexities that can give spurious results can be significantly minimised with an in-depth diagnostic tests. Therefore, researchers need to be aware of the estimation complexities due to the inherent characteristics of their data and be able to resolve these complexities to have appropriate estimation technique. This is more important when dealing with data that can be influence with individuals and households characteristics.

## 5.3. Research limitation and future works

The limitation of this study is that the modelling approach was applied to households' health data. The modelling approach may give a different result if apply to other households data such as consumption expenditure or income data. Therefore, future research can apply the modelling approach to other households' data to provide evidence of the relationship that are generalisable.

## References


Bajari, P., Hong, H., Khwaja, A., & Marsh, C. (2006). *Moral hazard, adverse selection, and health expenditures: A semi-parametric analysis* (Working Paper No. 12445). Cambridge, MA: NBER.

Baum, C. F., Schaffer, M. E., & Stillman, S. (2003). Instrumental variables and GMM: estimation and testing. *The Stata Journal, 3*(1), 1-31.

Bolhaar, J., Lindeboom, M., & van der Klaauw, B. (2008). *A dynamic analysis of the demand for health insurance and health care* (Discussion Paper, No. 3698). Bonn: IZA.

Cameron, A. C., Trivedi, P. K., Milne F., & Piggott, J. (1988). A microeconometric model of the demand for health care and health insurance in Australia. *Review of Economic Studies, 55*(1), 85-106.

Cutler, D. M., & Zeckhauser, R. J. (2000). The anatomy of health insurance. In J. A. Culyer & J. P. Newhouse (Eds.), *Handbook of Health Economics* (pp. 563-643). Amsterdam: Elsevier.

Hidayat, B., & Pokhrel, S. (2010). The selection of an appropriate count data model for modelling health insurance and health care demand: Case of Indonesia. *International Journal of Environmental Research and Public Health, 7*, 9-27. doi: 10.3390/ijerph7010009

Jones, A. M. (2000). Health econometrics. In J. A. Culyer & J. P. Newhouse (Eds.), *Handbook of health economics 1A* (pp. 265-344). Amsterdam: Elsevier.

Koç, Ç. (2005). Health-specific moral hazard effects. *Southern Economic Journal*, *72*(1), 98-118. doi: 10.2307/20062096

Liu, X., Nestic, D., & Vukina, T. (2012). Estimating adverse selection and moral hazard effects with hospital invoices data in a government-controlled health care system. *Health Economics, 21*(8), 883-901. doi: 10.1002/hec.1756

Newhouse, J. P., Sloss, E. M., Manning, W. G., & Keeler, E. B. (1993). Risk adjustment for a children's capitation rate. *Health Care Financing Review, 15*(1), 39-54.

Pagan, A. R., & Hall, D. (1983). Diagnostic tests as residual analysis. *Econometric Review, 2*, 159-218. doi: 10.1080/07311768308800039

Shea, J. (1997). Instrumental relevance in multivariate linear models: A simple measure. *The Review of Economics and Statistics, 79*(2), 348-352. doi: 10.1162/rest.1997.79.2.348

Staiger, D., & Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica, 65*(3), 557-586. doi: 10.2307/2171753

Vera-Hernández, A. M. (1999). Duplicate coverage and demand for health care: The case of Catalonia. *Health Economics, 8*(7), 579-598.

Waters, H. R. (1999). Measuring the impact of health insurance with a correction for selection bias: A case study of Ecuador. *Health Economics, 8*(5), 473-483.

Windmeijer, F. A. G., & Santos-Silva, J. M. C. (1997). Endogeneity in count data models: An application to demand for health care. *Journal of Applied Econometrics, 12*, 281-294. doi:10.1002/(SICI)1099-1255(199705)12:3<281::AID-JAE436>3.0.CO;2-1

Zweifel, P., & Manning, W. G. (2000). Moral hazard and consumer incentives in health care. In J. A. Culyer & J. P. Newhouse (Eds.), *Handbook of Health Economics* (pp. 409-459). Amsterdam: Elsevier.

***Appendix I: Description of the Variables used in the Analysis***

| Variable | Definition | Description |
|---|---|---|
| **Dependent Variables** | | |
| HINSTATUS | Health Insurance Status: Insured = 1, Non-Insured = 0 | Dichotomous |
| HINSTYPE | Health Ins Type: NHIS = 1, others = 0; Personal Health Insurance = 1, others = 0 | Dichotomous |
| MEDICONSUMP (*m*) | Number of times of Consuming Health Care Services | Count |
| **Independent Variables** | | |
| Married | Marital Status: Single = 1, Married = 2, Divorce/Separated = 3, Widowed = 4 | Categorical |
| Male | Gender Variable: Male = 1, 0 otherwise | Dichotomous |
| Age | The age of the respondent as at the last birthday | Continuous |
| FMTYPE | Family Type: Monogamy = 1, Polygamy = 2 | Categorical |
| FMHEAD | Head of the Family: Father = 1, Mother = 2 | Categorical |
| FMHEDUC | Head of the Family Level of Education: No formal Schooling = 1, Primary Education = 2, Secondary Education = 3, Post-Secondary Education = 4 | Categorical |
| SFMHEDUC | Spouse of the Family Head Level of Education: 1 = No formal schooling, 2 = Primary education, 3 = Secondary education, 4 = Post-secondary education | Categorical |
| FMHOCC | Head of the Family Occupation: Government Worker = 1, Formal Private Sector Worker = 2, Trader = 3, Transporter = 4, Farmer = 5, Self-Employed = 6, Housewife = 7, Unemployed = 8, others = 9. | Categorical |
| SFMHOCC | Spouse of the Family Head Occupation: Government Worker = 1, Formal Private Sector Worker = 2, Trader = 3, Transporter = 4, Farmer = 5, Self-Employed = 6, Housewife = 7, Unemployed = 8, others = 9. | Categorical |
| MEXPFD | Monthly Expenditure on Food | Continuous |
| MEXPTC | Monthly Expenditure on Transport & Communication | Continuous |
| MEXPHLT | Monthly Expenditure on Health | Continuous |
| MEXPORS | Monthly Expenditure on Others | Continuous |
| MTOTAEXP | Monthly Total Expenditure | Continuous |
| GHSTATUS | General Health Status measured using twelve questions about general well-being where high score indicates bad health status. | Continuous |
| COINS | Co-insurance Rate Paid by the insured | Continuous |
| PRICEHC | Price of Health Care Computed as Coinsurance Rate Multiply by Health Exp. | Continuous |
| PLACEACESS | Place of Access Health Care Facility: Self-Treatment = 1, Traditional Healers = 2, Private Hospital = 3, Government Hospital = 4, Pharmacy/Drug Shop = 5, Spiritual Home = 6, others = 7 | Dichotomous |
| HINSTYPE | Health Insurance Type: NHIS = 1, Private Company Health Insurance = 2, Personal Health Insurance = 3 | Dichotomous |
| MINCEMPL | Individual Monthly Income from Employment | Continuous |
| MINCGIFTS | Individual Monthly Income from Gifts | Continuous |
| MINCORS | Individual Monthly Income from Others | Continuous |
| MTINCO | Total individual monthly Income | Continuous |
| SICKINC | Income During Sick Period | Continuous |