Comparing the Parametric and Semiparametric Logit Models: Household Poverty in Turkey

Ebru CAGLAYAN

Corresponding Author: Department of Econometrics, Marmara University

Ressam Namik Ismail Sok. No.1, 34720, Bahçelievler, Istanbul, Turkey

Tel: 90-212-507-9925 ext.1236   Fax:90-212-539-1182   E-mail: ecaglayan@marmara.edu.tr

Tugba DAYIOGLU

Phd Student, Marmara University

Institute of Social Sciences, Istanbul,Turkey

E-mail: tugbaaksu@hotmail.com

Abstract

The objectives of this paper are to determine the factors which could affect the poverty status and living standards of a household and to illustrate the probabilities of poverty in households in Turkey using parametric and semi-parametric logit models. We use the data of the Turkish Household Budget Survey prepared by the Turkish Statistical Institute (TURKSTAT) for the year 2008. The semi-parametric method that combines the best features of the parametric and the non parametric approaches is introduced when the parametric model assumptions are violated. The results indicate that the most important determinants of poverty are the working status and occupation of household head, income, and ratio of worker in household and region.

1. Introduction

Poverty, a complex, multidimensional, and universal problem, has been conceptualized as income and material deprivation. Poverty shows differences from country to country and period to period, depending upon the developments in the level of welfare. Comparing countries, or periods within periods, with regard to poverty requires deciding who to call poor in the total population. The basic approach in the analysis of poverty is the detection of the poverty line. However, there may be varieties in detecting this line. This is because there must be a description of the poverty line when the gratitude of a household changes. In this sense, it is more suitable to define the poverty line of each society according to its own economic and cultural conditions (Kumar et al., 1996). Although there is no condition of extreme poverty in Turkey which is in the category of developing countries and within the group of middle income countries, poverty is still regarded as an important problem. The Turkish Statistics Institute (TURKSTAT) has released the results from its 2008 Household Budget Survey, revealing that 17.11 percent of the population lived below the poverty line in 2008, or 11.9 million people. According to TURKSTAT, none of Turkey's population lives on under $1 a day, the extreme poverty line. The number of individuals living below TURKSTAT's other established poverty lines of $2.15 and $4.3 was 0.47 percent and 6.83 percent, respectively. The percentage of people living below the hunger line, which takes into account only food expenditures, set at TL 275 per month for a four-person household, increased from 0.48 percent in 2007 to 0.54 percent in 2008. The report revealed that as the number of people in a household increased, the probability of living under the poverty line also increased. A full 26.95 percent of individuals living in urban households consisting of more than seven individuals lived under the poverty line, whereas this figure increased to a staggering 54.03 percent in rural areas. As a result of the effects of the global crisis household poverty increased.

The aims of this paper are to determine the factors which could affect poverty status and the living standards of households and to illustrate the probabilities of household poverty in Turkey. For these aims we estimate the parametric logit models against their semi-parametric alternatives. The reason for estimating a semi-parametric model apart from parametric logit models is to detect whether the effects of factors on poverty is parametric or non-parametric. We use the data of the Turkish Household Budget Survey prepared by the Turkish Statistical Institute for the year 2008.

The rest of the paper is organized as follows: The following section includes the introduction. Section 2 introduces the parametric and semi-parametric logit models. Sections 3 and 4 present data and empirical findings, respectively. The final section provides conclusions.

2. Parametric and Semi-parametric Logit Models

The logit model remains the most widely used parametric method for the estimation of binary dependent variable. This model depends on two assumptions: a known index which is assumed to influence choice, and a known parametric form for a distribution function which is assumed to yield choice probabilities. A Binary Dependent variable model has been used in this study. This class of models, dependent variable, may take on only two values zero and one. The binary logit model remains the most widely used parametric method for the estimation of binary choice models. The traditional parametric logit model approach to modelling binary choice is as follows:

$$E(Y|X = x) = p(Y = 1|X) = F(X^T\beta) = \frac{\exp(X^T\beta)}{1+\exp(X^T\beta)} \tag{1}$$

In the parametric approach to modelling, the function F is known and the values of the parameters are unknown. The estimation problem is to estimate the unknown parameters. The logit model is typically estimated by maximum likelihood. F is the cumulative logistic distribution function. The logistic distributions are symmetrical around zero. The fitted models can easily be interpreted and estimated accurately if the underlying assumptions are correct. If, however, they are violated then parametric estimates may be inconsistent and give a misleading picture of the regression relationship. Parametric models are typically chosen due to their tractability and ease of interpretation. However, the exact form of the response curve is usually unknown and even very complicated, so it is likely that the true model does not follow the logit model. If the functional form is mis-specified, then the estimates of the coefficients and the inferences based on them can be highly misleading. It is possible to relax the restrictive assumption that the functional form is known by using either semi-parametric or nonparametric models. In these types of models, the functional form is unknown. The problems of estimating semi-parametric and nonparametric binary response models have generated considerable interest in recent years.

There exists a rich and very impressive variety of approaches towards the semi-parametric estimation of binary response models including that of Coslett (1983), Ichimura (1986), Manski (1986), Rudd (1986), Ichimura and Lee (1991), Coslett (1991), Klein and Spady (1993), Lee (1995), Chen and Randall (1997), Picone and Butler (2000), and Ichimura and Thompson (1998) among others. Pagan and Ullah (1999) and Hardle and Horowitz (1996) also give a recent survey of semi-parametric approaches to the estimation of binary response models.

The semi-parametric logit model corresponds to a parametric logit model and is considered by generalizing the linear argument $(X^T\beta)$ to a partial linear argument $(X^T\beta + m(Z))$. The model expression is given as follows:

$$E(Y|X = x) = p(Y = 1|X, Z) = G(X^T\beta + m(Z)) = \frac{\exp(X^T\beta+m(Z))}{1+\exp(X^T\beta+m(Z))} \tag{2}$$

Where G (.) is a known function, is an unknown parameter vector, and m (.) is an unknown function. This model allows for the modelling of the influence of a part Z of explanatory variables in a nonparametric way. The parametric component and the nonparametric function m (.) can be estimated by the quasi likelihood method proposed in Severini and Staniswalis (1994).

3. Data

In this study we investigate the relationship expenditures per equivalent individual consumption, disposable income, social security coverage, working status (full time and part time), and percentage of household employer and ownership of housing using parametric and semiparametric logit models. Data used in the analysis is obtained from the Turkish Statistics Institute's 2008 Household Budget Survey on 17 500 sample households from the period 1 January 2008 to 31 December 2008.

In most studies the probability of household poverty is based on the consumption expenditures per equivalent individual. This is because the consumption of gender of the household head is different from each other. However, men have higher rates of consumption in a household than women. Another thing is that children have the least consumption in households. Consumption expenditures per equivalent individual have to include these experiences to show how many adult persons equal the household. TURKSTAT has done poverty calculations based on expenditure which is used with various methods. One of them is "relative poverty" and it is defined as the state in which the individual is below the average welfare level of the society. In this respect, those households having incomes and expenditures below a specified line compared to the general population shall be defined to be poor in a relative sense. As a welfare measure, consumption or income level may be selected according to the situation. In the Household Budget Survey, 50% of the median value of the consumption expenditures per equivalent individual is

defined as the relative poverty line and in this way the relative poverty rate is calculated. For this reason, we use consumption expenditures per equivalent individual to determine the dependent variable.

In this study, the expenditures per equivalent individual consumption (EPEIC) formulation is obtained in accordance with the OECD per equivalence scale (OECD, 2008) is calculated;

$$EPEIC = \frac{EHC}{MAPEI} \tag{3}$$

where EHC and MAPEI refer to the expenditures of household consumption and magnitude of adult per equivalent individual, respectively. Magnitude of adult per equivalent individual is calculated as follows:

MAPEI= [1+ (0.5*(the number of persons older than 14 years old-1) +0.3*(the number of persons younger than 14 years old))]

The OECD equivalence scale assigns a value of 1 to the first household member, of 0.7 to each additional adult and of 0.5 to each child. This scale (also called the Oxford Scale) was mentioned by the OECD (1982) for possible use in "countries which have not established their equivalence scale''. For this reason, this scale is sometimes labelled the OECD scale (OECD, 1982). In the Household Budget Survey, 50% of the median value of the consumption expenditures per equivalent individual is defined as the Relative Poverty Line. Following all these calculation, the dependent variable is determined as:

If the household consumption expenditures < relative poverty line then Y=1

If the household consumption expenditures > relative poverty line then Y=0

The explanatory variables we use are: disposable income (DISINC), social security coverage (SSC), working status (WORKINGST), ratio of worker in household (RAW), owning of housing (OWNH), age of the household (AGE), region (REG), gender (GEN) and occupation of the household (OCCU). The construction of the explanatory variables has been explained in the Table 1.

[Insert Table 1 here]

4. Empirical Findings

In this study we investigated the effect of the crisis on the probability of household poverty using Semiparametric Logit models. Then we tried to find the best model. So we compared the parametric logit model, nonparametric and semiparametric logit models. We used the comparing criteria to find the fitted model. According to the variables in the above, the parametric logit, nonparametric and semi-parametric logit summary of the results obtained were as follows:

In parametric logit models, the coefficients of disposable income, working status, percentage of working, regular wage worker, jobber and region are statistically significant at the 1% level. However, the coefficient of the social security, age, sexuality are found statistically insignificant, thus we ignore them in the model. Model I shows the results of the parametric logit model. These results show that the entire parameter coefficient is statistically significant except age and social security coverage. Models II, III and IV are semi-parametric logit models. Along with the variables of income and number of workers in Model II, only income variable in Model III and only number of workers variable in Model IV were estimated non-parametrically. Since other variables are dummy variables, the models were estimated parametrically. Estimation results of all models are summarized in Table 2. The coefficients of all the models in the table are statistically significant.

[Insert Table 2 here]

In this study to compare the performances of the parametric and semiparametric logit models, deviance residuals and graphics of absolute deviance residuals were examined.

Absolute deviance residuals examined in the detection of goodness of model can reveal the features of residual points. Therefore, absolute deviance residuals are especially suggested as much as deviance residuals which show the goodness of model (Littel et al., 2002).

First of all, the graphics of the parametric logit model are given in Graphic 1. Secondly household percentage in the semi-parametric logit model and graphic of the nonparametric model of disposable income are given in Graphic 2. Thirdly, the graphics of nonparametric model of income are given in Graphic 3. And finally, the graphics of the nonparametric models of household working percentage are given in Graphic 4.

[Insert Graphic 1 here]

When we observe the graphics above, 1(a) shows the model goodness of the parametric logic model. The exponential distribution was chosen from binomial distribution. For the model goodness of the logit model,

alternatives of binomial distribution are chosen and logit models are considered (Härdle and Horowitz, 1996; Dunn and Smyth, 1996; Goegebeur, 2008). In this graphic, variables of disposable income, household working percentage, regular wage worker, casual worker, and region were significant and therefore considered. Data points of variables were ordered in bold line. In Graphic 1(b) and Graphic 1(c) deviance residuals and absolute deviance residuals were considered. Graphics of deviance residuals are generally parallel. Values of deviance residuals are generally located on direct distinctive curves. The direct line in the middle is the normal distribution curve. When there is normal distribution these distributions are located on line. Reaction variables show possible values. Lines formed by bold black bubbles in the deviance residuals graphic show the deviance residual of parametric logit model. Deviance residuals in the mentioned graphic continue symmetrically around the zero line. This situation is similar to the appearance of ordinary residuals, and two bold black lines at the top and bottom move towards the zero line, the direct line not being in ordinary residuals but in curvilinear form. In absolute deviance residual deviation curve does not cross any line. Both absolute deviance residual and deviance residual move on a linear curve.

In the Semi-parametric Logit Model, Graphic 2(a) shows model goodness. When this graphic is considered, it is seen that the semi-parametric model is not on a parametric logit curve and the line formed by bold black bubbles moves on a different line. It is observed that these two lines are different from each other. While the direct line shows parametric structure, lines formed by both top and bottom black bubbles show the semi-parametric logit model. The previously mentioned bold line is far from the linear line in the middle.

When deviance residuals in Graphic 2(b) are observed, the semi-parametric logit model gets over horizontal appearance and shows curvilinear progress. Residual deviance cuts the parametric logit model curve. However, this is not on the curve and progress curvilinear on horizontal axis. This situation shows a semi-parametric structure, not a parametric one. Residual deviance progresses symmetrically on a linear line both on and under the horizontal line (Note 1). In semi-parametric logit model structure this distribution inclines both towards the top and the bottom. Graphic 2(c) shows absolute residual deviance. The bold black bubbled line shows absolute residual deviance of the semi-parametric logit model. Absolute residual deviance progresses including the residual below and above the horizontal line and shows curvilinear structure rather than residual deviance. Absolute residual deviance is close to the semi-parametric logit model which is at the bottom of the graphic and far from the curve which is a direct line. It is far from having semi-parametric structure on this curve. However, the semi-parametric logit model progressing on the bold black line as black bubbled line shows that the structure reflects instead the semi-parametric logit model.

Graphic 2 (d and e) shows the smoothing parameters for Household Working percentage and Disposable Income which are the non-parametric variables in the semi-parametric logit model. The leveled variable is generally within or on the disposable line. Mentioned variables are smoother between the two determined bands (Note 2). This situation shows that the variable is getting better condition. In the goodness of fit graphic of variable linear, the trend is not erased and variables are levelled on a curvilinear form. In this condition, the levelled disposable income variable has non-parametric structure due to the goodness of fit.

[Insert Graphic 2 here]

In this semi-parametric logit model, only the disposable income variable was used in non-parametric structure. In this model, the self-employed (OCCU3) and free- family worker (OCCU4) variables were found meaningful and therefore was included in the model. When Graphic 3(a) was considered, goodness of fit was observed and the black bold line belongs to the semi-parametric logit model. Direct line shows parametric logit structure and is close to the linear trend line. However, the bold black line which is in the semi-parametric logit model is observed both above and below in the direction of logit values which are dependent variables. Much as the two lines separate from each other, they are far from a parametric structure.

Graphic 3(b) shows residual deviance. Here the bold black line moves away from the parametric curve and shows semi-parametric logit model residual deviance. While the black bold line progresses on a linear structure in residual deviance progress, residual deviances in the semi-parametric logit model graphic are seen as a separate line on a semi-parametric logit model graphic. This shows a non-parametric structure.

[Insert Graphic 3 here]

Graphic 3(c) shows an absolute deviance residual and includes unobserved residuals too, and it has a more curvilinear structure than residual deviance and the bold black line is far from the direct line which shows parametric structure. This situation shows that the semi-parametric logit model is a better model. Graphic 3(d) is the smoothing parameter of disposable income variable which has a non-parametric structure. When this graphic is observed, it is seen that disposable income is smoothed with splay smoothing and a better semi-parametric logit model is obtained by using a non-parametric structure. The black bubbled line between two bands in the graphic is grouped on a direct line. In other words, the variables are ordered. In this sense, it is a fact that only the income variable should be used in the non-parametric structure.

When the graphics are observed, in the semi-parametric logit model in which household working percentage is not parametric, Graphic 4(a) is the conformity graphic of this model and the bold black line shows the values of a semi-parametric logit model. Parametric logit values are on a thin black line. This line is close to the line which shows a cross linear trend. The previously mentioned parametric and semi-parametric curves are far from each other. This situation shows that the structure is not parametric. Graphic 4(b) shows residual deviances. The residual deviances of this model have taken high values. Residual deviance is again curvilinear and has a form which decreases below zero but increases above zero. In other words, deviation is at high levels. Graphic 4(c) shows the absolute residual deviances and these values are positive. These values show us the values of ignored residuals. In this graphic, disseminated absolute residuals have a bold black line that increases towards the top. They have a curvilinear progress. The direct black line shows parametric structure, but in the absolute residual deviance graphic gaps between these lines are quite apparent and this shows us that the structure is not parametric. Graphic 4(d) shows a smoothing parameter. Here the variable was smoothed with splay smoothing, however including this variable in the model does not give a good result. In the previously mentioned graphic, the variable which should be between suitable bands is located on a direct line. This situation shows that the variable being included in the model by being smoothed is not enough. The afore-mentioned variable is not adequate in a non-parametric structure. However, it has given a better result than the parametric logit model.

[Insert Graphic 4 here]

When all the semi-parametric logit models are considered separately for 2008, it is seen that the semi-parametric logit model has the best performance among others. Therefore, coefficients of this model are interpreted and findings about poverty will be evaluated.

Coefficients are interpreted directly in the semiparametric logit models as in the parametric logit model and odds ratios are calculated and comments are made according to these values. Odds ratio give the probability of the emergence of poverty coded as "1" in the dependent variable. Odds ratio are interpreted differently for continuous variables. The odds ratios of continuous variables are calculated by imposing quota or by making the variable dashed (Kemp, 2000; Nash and Bradford, 2001; Stacey and Tatum, 1985). First of all odds ratios were calculated for the variables in parametric part of semi-parametric logit model in which both the variables of household and disposable income are parametric. The odds ratio for regular working was found to be exp (-1.4429) = 0.2362. In order to make the comment more logical, if the values are negated the value of 4.347826 is obtained. In this sense indicator and reference values shift place. This means the poverty probability of employers, self-workers, casual workers and free family workers is 4.3478 times more than regular-waged workers. The odds rate for casual workers was found to be exp (1.5501) =4.712110. According to this, it can be said that the poverty probability of casual workers is 4.71 times more than regular-waged workers, employers, self-employed workers and free-family workers. The odds rate was found to be exp (6.9272) = 1019.63 for part-time workers and as exp (6.1323) = 460.514 for the region. According to these results, it can be said that the poverty probability of part-time workers is higher than full-time workers. The poverty probability of rural workers is higher than urban workers.

In the semi-parametric logit model which shows the best performance, the variables (income and number of workers in household) which form the non-parametric part are continuous variables. In this sense, the odds ratios were calculated differently compared with other variables (Note 3). The income variable was taken as the minimum wages of 2008. The minimum wage of this year is 435.92 TL. In this frame, the income variable was sectioned as lower and higher than the minimum wage and the odds rate is exp (3.12630253) = 22.789559. In this sense, those whose income is lower than the minimum wage have more probability of having poverty compared to those whose income is higher. The odds rate for the variable of working people percentage is based on 50 % of workers in the household and this variable was sectioned as those lower and higher than 50% (Note 4). In this sense, the odds rate was calculated as exp (0.00894947) = 1.008989. The poverty probability of those who work less than 50% is 1.0089 times more than those who work more than 50%.

When we evaluate the findings generally, we have proof that the poverty probability of part-time workers is higher than full-time workers, rural workers over urban workers, casual workers over regular-waged, employers and self-employed workers. While the effect of these variables on poverty is explained parametrically, it was found that the effect of income and number of workers were non-parametric.

5. Conclusions

The "Household Budget Survey" is one of the major sources providing information on the expenditures patterns, living standards and income levels of the households by socio-economic groups and urban-rural settlements and regions. This publication has been prepared to provide users with the main indicators, related to the consumption expenditures of the households for the whole of Turkey as well as urban and rural settlements, gathered from the Household Budget Survey conducted on 17,500 sample households for the period 2008.

In order to determine the most suitable model in the study and detect whether the effect of the variable on poverty is parametric or nonparametric, parametric logit and semi-parametric logit models were estimated and results were compared. Proof was obtained in the way that the model which defines the factors best is the semi-parametric model and variables of income and number of household workers were to be found in the model non-parametrically.

When poverty is considered in the urban and rural sense, it was observed that poverty was higher in rural areas and the situation was confirmed with the reports of TURKSTAT. In the poverty research of TURKSTAT, it is seen that the rate of poverty in the rural section is on a large scale. Therefore, necessary regulations should be conducted in order to overcome the poverty in rural areas. Moreover, the reasons for participating / not participating in the labour force should be observed and policies should be developed in order to overcome these reasons. First of all, people who are at or just below the poverty line should be raised above this limit. Policies for strengthening the condition of non-workers should be handled with long-term precautions. In taking all these steps, detecting the deficiencies which are current in the scope of struggling with poverty, and making the necessary changes in this sense and forming policies, would be effective ways of decreasing the poverty of the country.

References

Chen, H. & Randall, A. (1997). Semi-nonparametric estimation of binary response models with an application to natural resource valuation. *Journal of Econometrics*. 76, 323–340. doi:10.1016/0304-4076(95)01794-1, http://dx.doi.org/10.1016/0304-4076(95)01794-1

Coslett, S.R. (1983). Distrubition–Free Maximum Likelihood Estimation of the Binary Choice Mode. *Econometrica*. 51, 765-782. doi:10.2307/1912157, http://dx.doi.org/10.2307/1912157

Coslett, S. (1991). Semiparametric Estimation of A Regression Model With Sampling Selectivity. In W. Barnett, J. Powell & G. Tauchen (Eds.). Nonparametric and Semiparametric Methods in Econometrics and Statistics. Cambridge University Press, Cambridge.

Dunn, P.K. & Smyth G.K. (1996). Randomized Quantile Residuals. *Journal of Computational and Graphical Statistics*. 5(3), 236-244. doi:10.2307/1390802, http://dx.doi.org/10.2307/1390802

Fox, J. (2002). Nonparametric Regression. In Appendix to An R and S-PLUS Companion to Applied Regression. [Online] Available: http://cran.r-project.org/doc/contrib/Fox-Companion/appendix.html. (March 10, 2010)

Gharibvand L. & Fernandez G. (2007). Survival Analysis Plots Using SAS, ODS Graphics. University of Nevado, Reno. Available: www.crda.unr.edu/crda/publications/publications.htm. (June 24, 2010)

Goegebeur, Y. (2008). Logistic Regression,Poisson Regression and Generalized Linear Models. Deparment of Statistics University of Southern, Denmark. Available: http://www.stat.sdu.dk/matstat/yuri-st504/w6slides2.pdf. (January 16,2011)

Härdle, W. & Horowitz, J. (1996). Direct Semiparametric Estimation of Single-Index Models With Discrete Covariates. Journal of the American Statistical Association. 91, 1632-1640. doi:10.2307/2291590, http://dx.doi.org/10.2307/2291590

Ichimura, H. & Lee L. (1991). Semiparametric Least Squares Estimation Of Multiple İndex Models: Single Equation Estimation. In W. Barnett, J. Powell & G. Tauchen (Eds.). Nonparametric and Semiparametric Methods in Econometrics and Statistics. Cambridge University Press, Cambridge.

Ichimura, H. (1986). Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single Index Models. *Journal of Econometrics*. 58, 71-120. doi:10.1016/0304-4076(93)90114-K, http://dx.doi.org/10.1016/0304-4076(93)90114-K

Ichimura, H. & Thompson, T. (1998). Maximum Likelihood Estimation of A Binary Choice Model With Random Coefficients Of Unknown Distribution. *Journal of Econometrics*, 86, 269–295. doi:10.1016/S0304-4076(97)00117-6, http://dx.doi.org/10.1016/S0304-4076(97)00117-6

Kemp G.C.R. (2000). Semiparametric Estimation of a Logit Model. University of Essex, Available: http://privatewww.essex.ac.uk/~kempgcr/papers/logit_mx.pdf. (January 16,2011)

Klein, R. W. & Spady, R. H. (1993). An Efficient Semiparametric Estimator for Binary Response Models. *Econometrica*. 61 (2), 387–421. doi:10.2307/2951556, http://dx.doi.org/10.2307/2951556

Kumar,T.K., Göre, A.P. & Sitaramam,V. (1996). Some Conceptual and Satatistical Issues on Measurement fo Poverty. *Journal of Statistical Planning and Inference*. 49(1), 53-57. doi:10.1016/0378-3758(95)00030-5, http://dx.doi.org/10.1016/0378-3758(95)00030-5

Lee, L.F. (1995). Semiparametric maximum likelihood estimation of polychotomous and sequential choice models. *Journal of Econometrics.* 65, 385-428. doi:10.1016/0304-4076(93)01591-9, http://dx.doi.org/10.1016/0304-4076(93)01591-9

Littel, R.C., Stroup, W.W. & Freund, R.F. (2002). SAS for Linear Models. (4th ed.). SAS Institute Inc. North Carolina: USA.

Manski, C.F. (1986). Semiparametric Analysis Of Binary Response From Response-Based Samples. *Journal of Econometrics.* 31(1), 31-40. doi:10.1016/0304-4076(86)90054-0, http://dx.doi.org/10.1016/0304-4076(86)90054-0

Nash, M.S. & Bradford, D.F. (2001). Parametric And Nonparametric Logistic Regressions For Predictions Of Presence/ Ansence of An Amphibian. US Enviromental Protection Agency Office of Research and Development, EPA/600/R-01/081, 1-35. Available: http://epa.gov/esd/land-sci/pdf/008leb02.pdf. (December 20,2010)

OECD (1982). The OECD List of Social Indicator. Paris.

OECD (2008). Growing Unequal ? Income Distribution and Poverty in OECD Countries. Available: http://www.oecd.org/dataoecd/45/42/41527936.pdf. (January, 2 2011)

Pagan, A. & Ullah, A. (1999). *Nonparametric Econometrics*. Cambridge University Press, Cambridge.

Picone, G.A & Buttler J.S. (2000). Semiparametric Estimation of Multiple Equation Models. *Econometric Theory*. 16, 551-575. doi:10.1017/S0266466600164047, http://dx.doi.org/10.1017/S0266466600164047

Ruud, P. (1986). Consistent Estimation of Limited Dependent Variable Models Despite Misspecification of Distribution. *Journal of Econometrics*. 32, 157–187. doi:10.1016/0304-4076(86)90017-5, http://dx.doi.org/10.1016/0304-4076(86)90017-5

Severini, T.A. & Staniswalis, J.G. (1994). Quasi-likelihood Estimation in Semiparametric Models. *Journal of the American Statistical Association*. 89(426), 501-511. doi:10.2307/2290852, http://dx.doi.org/10.2307/2290852

Stacey C.I. & Tatum T. (1985). House Treatment With Organochlorine Pesticides and Their Levels in Human Milk –Perth,*Western Australia. Bulletin of Environmental Contamination and Toxicology*. 35, 202-208. doi:10.1007/BF01636499, http://dx.doi.org/10.1007/BF01636499

Notes

Note 1. Please look for similar explanation : Gharibvand and Fernandez, 2007.

Note 2.Spline smoothing was conducted in this study. Variables which are applied spline smoothing became more smooth with degree of freedom. For this smoothing process look; Fox, 2002.

Note 3. If income < minimum wage (435.9 ytl in 2008) the income coded 1, If income > minimum wage income, the income coded as'' 0''

Note 4.If household working percantage is < %50, household working percantage is coded '1', others '0'.

Table 1. Description of Explanatory Variables

| Variable	Short name	Description
Disposable Income	DISINC	Turkish Liras
Region	REG	1 if rural, otherwise (urban) 0
Age of Household	AGE	AGE1=1 for less than 30 years, others=0 AGE2=1 for more than 30 years and less than 40 years, others=0 AGE3=1 for more than 40 years and less than 60 years, others=0
Social Security Coverage (BAGKUR, SSK, etc.)	SSC	1, if she/he has social security coverage, otherwise 0
Working Status	WORKINGST	1 if full time, otherwise (Part time) 0
Gender	GEN	1 if male, otherwise (female) 0
Ratio of worker in household	RAW	%
Occupation of the household	OCCU	OCCU1=1, if household belongs to category of regular wage worker, others=0 OCCU2=1, if household belongs to category of jobber, others=0 OCCU3=1, if household belongs to self-employed category of workers, others=0 OCCU4=1, if household belongs to free- family worker category, others=0

Table 2. Results of the Parametric and Semiparametric Logit Models.

Variables	Model I Parametric Logit	Model II Semiparametric Logit (DISINC and RAW nonparametric)	Model III Semiparametric Logit (DISINC nonparametric)	Model IV Semiparametric Logit (RAW nonparametric)
	Coefficient (standard error)			
OCCU1	-1.4301 (0.3905)	-1.4429 (0.3938)	-1.7803 (0.5368)	-1.6259 (0.5018)
OCCU2	1.4738 (0.2724)	1.5501* (0.2728)	2.8255 (0.4497)	4.2385 (1.8428)
OCCU3			-2.8915 (0.4432)	-1.4638 (0.5999)
OCCU4			8.3358 (3.3134)	
WORKINGST	1.3554 (0.4458)	6.9272 (2.2823)	2.7522 (0.2986)	2.9016 (0.9303)
REG	1.3811 (0.2665)	6.1323 (2.2445)	5.0670 (1.7754)	3.6022 (0.2960)
DISINC	2.1617 (0.6525)	**1.1170 (0.2696)**	**1.6952 (0.6579)**	3.1156 (0.6049)
RAW	1.1125 (0.2483)	**3.9613 (0.4462)**	3.3656 (1.4297)	**2.8017 (0.9161)**
CONSTANT	-1.2910 (0.0445)			
R^2	0.346	0.796	0.932	0.7855
F(parametrik)	76.8439	37.082	122.438	32.7718
F(nonparametrik)		184.46	698.6	56.2266
Residual Deviance	8928.89	1092.697	903.362	1428.804

Notes: Bold numbers refer to the smoothing parameter in the semiparametric logit model.

All coefficients are significant at the level of 1%.

Graphic 1: Parametric Logit Model
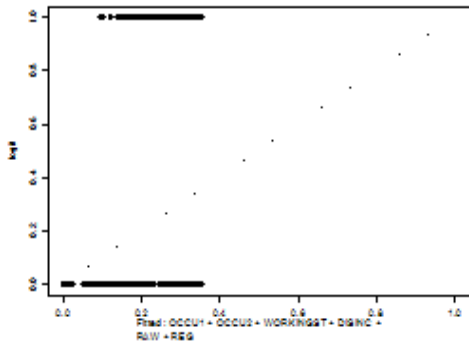
(a) parametric logit model   **(b) deviance residuals**
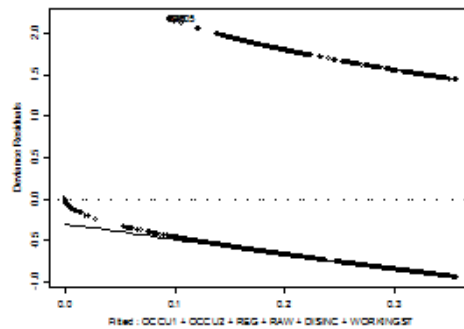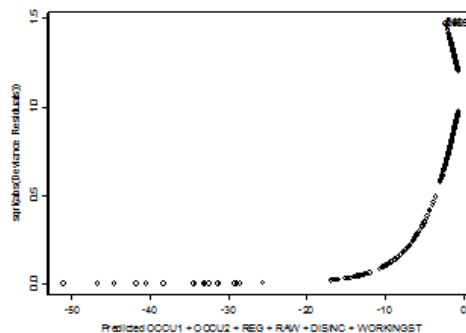
(c) absolute deviance   residuals

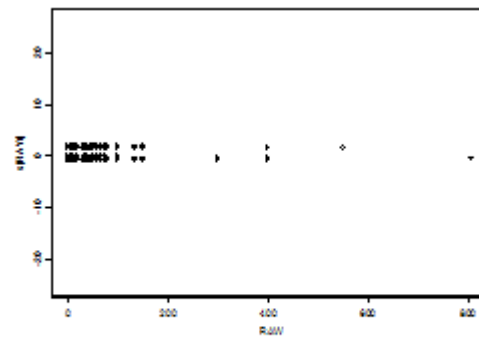**Graphic 2. Semi-parametric Logit Model (DISINC and RAW nonparametric)**
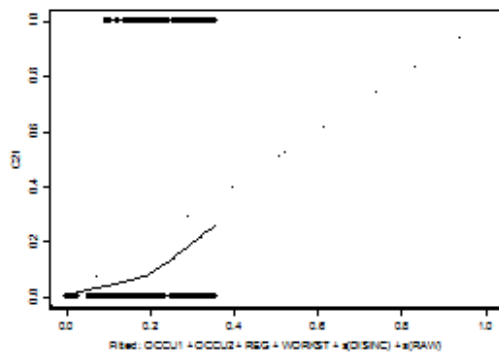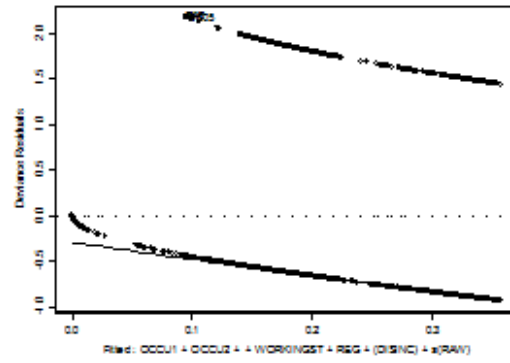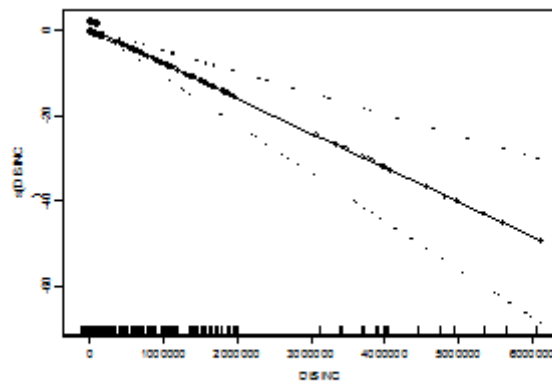

(a)

Semi- parametric logit model

(b)

deviance residuals

(c)

absolute deviance residuals
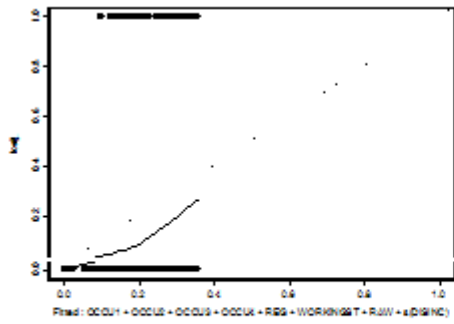
(d)
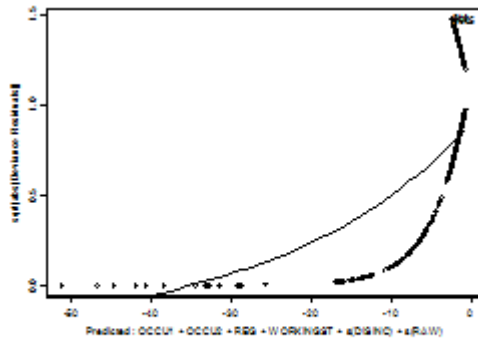
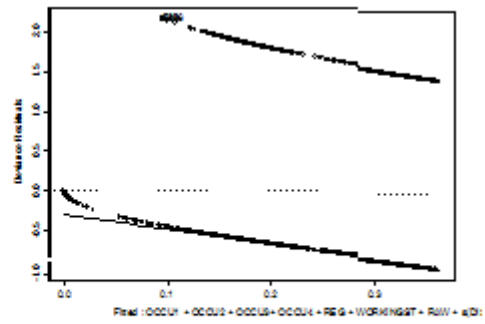**smoothing parameter of RAW**

(e)

smoothing parameter of DISINC
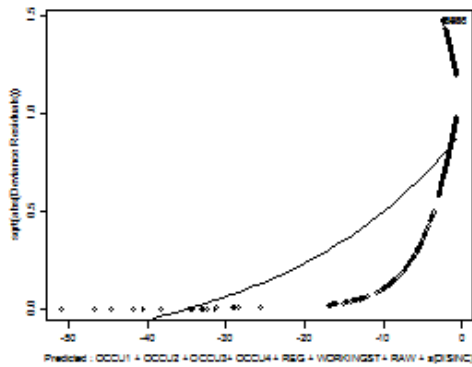
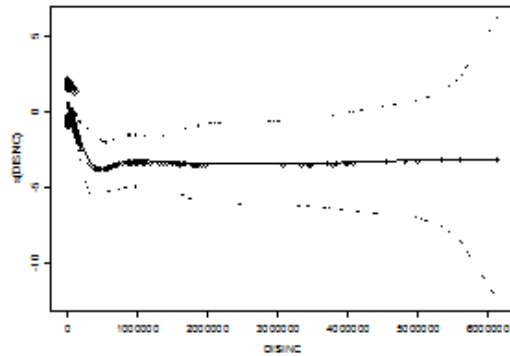**Graphic 3. Semiparametric Logit Model (DISINC nonparametric)**

(a)

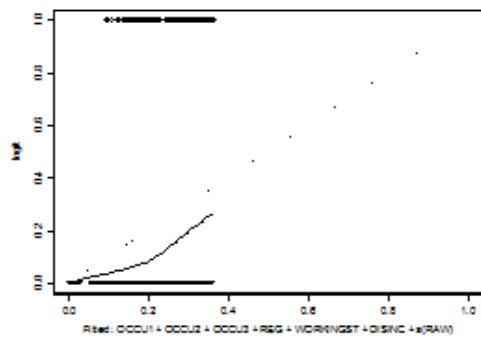**Semi- parametric logit model**

(b)

deviance residuals

(c)

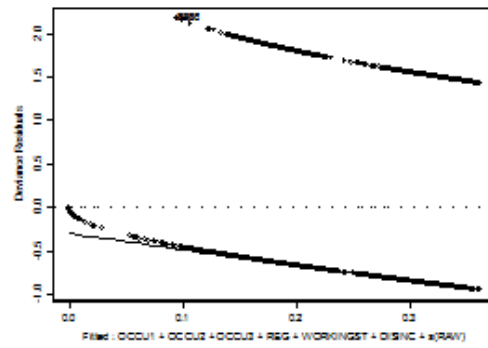**absolute deviance   residuals**

(d)

smoothing parameter of DISINC
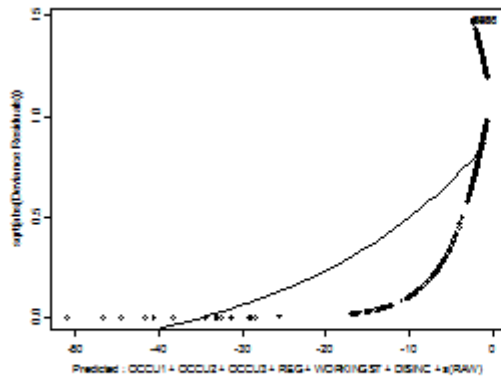
Graphic 4. Semi-parametric Logit Model (RAW nonparametric)
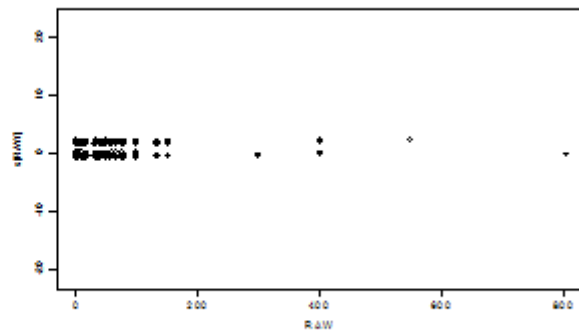

(a)

Semi- parametric logit model

(b)

deviance residuals

(c)

absolute deviance residuals

(d)

smoothing parameter of RAW