HOUSEHOLD WHOLE AND LOW-FAT MILK CONSUMPTION IN POLAND:

A CENSORED SYTEM APPROACH

by

SHENGFEI FU

(Under the Direction of Cheolwoo Park)

ABSTRACT

Milk products are a preeminent food category in Poland, providing both employment and dietary benefit. This thesis investigates factors affecting household milk consumption in Poland, paying attention to the effect of outmigration, an issue in Poland. We apply both the bivariate two-part model (B2P) and multivariate sample-selection model (MSSM) to the milk consumption data in Poland and choose the bivariate two-part model based on theoretical, practical, and statistical grounds. We estimate actual milk expenditure and explain the dependence of whole and low-fat milk consumption decisions using the B2P model.

INDEX WORDS: dairy product, milk consumption, whole milk, low-fat milk, Polish household, Poland, depopulation, worker migration, nutrition, dietary welfare, censored system, bivariate two-part model, Heckman’s sample-selection model, multivariate sample-selection model.

HOUSEHOLD WHOLE AND LOW-FAT MILK CONSUMPTION IN POLAND:
A CENSORED SYTEM APPROACH

by

SHENGFEI FU

B. A., University of International Relations, China, 2009

M. S., The University of Georgia, USA, 2012

A thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

STATISTICS

ATHENS, GEORGIA

2014

© 2014

Shengfei Fu

All Rights Reserved

HOUSEHOLD WHOLE AND LOW-FAT MILK CONSUMPTION IN POLAND:

A CENSORED SYSTEM APPROACH

by

SHENGFEI FU

Major Professor: Cheolwoo Park

Committee: Wojciech J. Florkowski

Jaxk Reeves

Electronic Version Approved:

Julie A. Coffield
Interim Dean of the Graduate School
The University of Georgia
December 2014

ACKNOWLEDGEMENTS

I would like to express my gratitude to everyone who supported me through the course of this thesis. I am thankful for their aspiring guidance, invaluably constructive criticism and friend advice.

I would like to thank Dr. Wojciech J. Florkowski, my advisor for my PhD program in Agricultural and Applied Economics, for his consent and support on my taking a secondary degree in Statistics. I am grateful to Dr. Florkowski for initiating this research topic and allowing me to use it as my thesis. I would also like to thank Dr. Florkowski for sharing his illuminating knowledge on consumer demand theory and the dairy sector in Poland.

I would like to express my thanks to Dr. Cheolwoo Park, my advisor in Statistics, for sharing constructive views and advice on various issues related to my thesis. I am thankful for his excellent guidance, patience, motivation, and being very approachable and helpful.

I would also like to thank Dr. Jaxk Reeve for serving as my committee member. The knowledge and experience I learned from his classes are constructive to the completion of this thesis.

And lastly, I would like to thank Dr. Anna Klepacka for collecting the Polish migration data used in this thesis and for sharing her knowledge on the dairy sector in Poland.

TABLE OF CONTENTS

Page

CHAPTER

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

Dairy products are a preeminent food category in the retail sector of many countries (Sznajder, 1999). Dairying is among the most important farm enterprises across Europe (Wilczyński, 2013). And in Poland, the dairy sector is one of the most important parts of the food industry (Sznajder, 2012). It represents about 16% of sales revenues of the Polish food processing industry (Sznajder, 2012).

Dairy products are also paid attention to in literature on food consumption and nutrition, because of their importance for disease prevention and health maintenance. In countries located in the temperate zone, milk has been a major source of essential nutrients including calcium, and vitamins D and A, among others. However, recent years has observed a substantial worldwide decline in fluid milk consumption, a trend that concerns nutritionists and health scientists.

Given the importance of the dairy production, processing and retailing, and milk's essential role in the diet of consumers in Poland, factors responsible for consumption deserve a closer scrutiny.

Food demand literature has identified a variety of socio-economic and demographic factors as consumption determinants, including household income, household size and structure, region of residence, and individual characteristics such as age, education level, and employment status. A special factor in Poland is worker migration and depopulation, especially after Poland's accession to the EU in 2004, coupled with free job market entry to other EU countries.

With macroeconomic development and demographic changes particularly associated with worker migration and depopulation, the dietary patterns are expected to have substantially changed. Migration leads to changes in age structure and gender composition, which in return contributes to different consumption features. The resulting difference in food consumption can contribute to insufficient or unbalanced nutrition intake and thus results in a less healthy population. The combination of relatively lower incomes, unfavorable population changes and dietary insufficiency creates conditions for the emergence of persistently underdeveloped areas.

Previous studies focused on the dampening effect of depopulation on economic growth; however, less attention has been paid at a micro/household level to the dietary welfare of people living in the depopulating regions. This thesis investigates factors affecting household milk consumption in Poland, accounting for the effect of outmigration.

Despite the attention paid to dairy products in literature on food consumption and nutrition, no study, to our best knowledge, has examined the consumption patterns in Poland. This paper contributes to fill this gap in literature.

The investigation of demographic, socio-economic and location factors, and their connection to milk consumption are important because of milk's dietary benefits. Furthermore, the declining fluid milk consumption over the years has forced the re-structuring of the dairy processing sector and affected local job opportunities in Poland.

Therefore, an analysis of factors influencing milk consumption, including milk of various fat contents, will offer insights applicable in milk processing and distribution, and, even in assessment of potential public health threats resulting from permanent decline in fluid milk consumption. Lastly, expanding milk production in regions with suitable natural conditions could provide job opportunities in rural areas (Klepacka et al., 2013). Insights about household milk consumption decisions are helpful to that effort.

The remainder of this thesis is organized as follows: Chapter 2 describes the methodology, including economic theory and statistical modeling. Chapter 3 introduces data source and variable definitions. Chapter 4 reports estimation results and goodness of fit. Finally, chapter 5 concludes with discussion.

CHAPTER 2

METHODOLOGY

This chapter describes economic theory, statistical modeling and relevant empirical tests for model choice.

2.1 Economic Theory

Researchers have long hypothesized a two-stage choice process where consumers first decide whether to buy a commodity, and then choose the amount to purchase (e.g. Bettman 1979; Gensch 1987; Shocker et al. 1991; Wright and Barbour 1977). This study follows that well-accepted hypothesis. These two stages of decision-making process are referred as the *participation decision* and *level decision*.

A qualitative choice model based on a random utility maximization developed by McFadden (1980) provides the theoretical foundation for model specification. A household maximizes the random utility function subject to a budget constraint. The household random utility function is given by:

$$V(y,q;\boldsymbol{w}) = d \cdot U(y,q;\boldsymbol{w}) + (1-d) \cdot U^{*}(q;\boldsymbol{w}) \qquad (1)$$

where U is the utility for buyers and U* for non-purchasers; y is the quantity of milk with price p, which implicitly enters the utility function through a budget constraint; q is a composite commodity for other goods with price normalized to one; **w** is a vector of demographic variables; and d is a binary variable that equals one if the household buys milk and zero otherwise.

Assume the outcome for milk purchase, the participation decision, is generated by a binary choice structure as follows, where for convenience, observation subscription is omitted:

$$d = \begin{cases} 1 \; if \; \boldsymbol{z}'\boldsymbol{\alpha} + u > 0 \\ 0 \; if \; \boldsymbol{z}'\boldsymbol{\alpha} + u \leq 0 \end{cases} \tag{2}$$

where **z** is the vector of variables affecting the binary purchase decision, while $\boldsymbol{\alpha}$ is the corresponding vector of parameters; u is a random error, usually assumed to be normally distributed in econometric analyses.

In cross-sectional demand modeling, zero observations are often treated as the result of economic non-consumption (i.e., corner solution in an optimization problem). In some cases, however, zero purchase might be caused by behavioral factors other than prices. Because y does not enter the purchasers' utility function $U^*(q; \boldsymbol{w})$ as described in equation (1) and p > 0, the optimal level is y = 0 for a non-eater. This optimal zero purchase could be a corner solution or the result of opting out of the market.

For a buyer, the optimal level of y results from a solution to the constrained utility maximization problem with a fixed budget I:

$$\max_{y,q} \; \{U(y, q; \boldsymbol{w}) | \; py + q = I\}. \tag{3}$$

Assume that the utility function $U(y, q; \boldsymbol{w})$ is regular strictly quasi-concave and has positive first partial derivatives with respect to y and q. Furthermore, assume an interior solution for y and q. Then, solving Equation (3) yields the notional (latent) demand for milk, y*.

Further assume the latent variable y^* is expressed by the lognormal distribution to ensure positive value and to mitigate right skewness:

$$\log(y^*) = \boldsymbol{x}'\boldsymbol{\beta} + v \qquad (4)$$

where $\boldsymbol{x}$ and $\boldsymbol{\beta}$ are variables and corresponding parameters affecting level decision, respectively; and v is a random error that follows normal distribution. Vector $\boldsymbol{x}$ usually has a wide range of common factors as vector $\mathbf{z}$ in Equation (2)**.** Sometimes they are identical.

Therefore, the participation equation and the level equation form a hurdle model, which constitutes two parts, one part generating zeros and the other part generating positive expenditure:

$$\mathrm{y} = \begin{cases} 0 \text{ if } \mathrm{d} = 0 \\ \mathrm{y}^* \text{ if } \mathrm{d} = 1 \end{cases} \qquad (5)$$

where the binary outcome d is governed by a discrete choice process, as described in Equation (2) and the latent milk expenditure y^* is assumed to be a lognormal distribution as in Equation (4).

2.2 Statistical Modeling

The occurrence of excessive percentage of zeros in micro-data sets mandates a proper treatment for the censoring of the dependent variables. Such zero observations may occur for three main reasons: infrequency of purchase in survey data with short recording periods, some individuals are out of market for various reasons (for example, lactose intolerance in the case of milk consumption), and economic non-consumption under current price and individual income. The particular interpretation given to zero observations can have a crucial bearing on the estimation approach adopted (Madden 2008). Various modeling structures are proposed in the existing literature to accommodate censored data, including the Tobit model (Tobin 1958), two-part model (Cragg 1971), and Heckman's sample selection model (Heckman 1976, 1979).

The Tobit model uses the same set of parameters to determine the probability and level of purchase. Furthermore, it implicitly assumes that an explanatory variable's effect has the same direction in both data-generating processes. Cragg (1971) first discusses a more flexible form of Tobit model, which allows two different sets of parameters, one to determine the probability of a limit observation and the other set the density of the observations. Lin and Schmidt (1984) propose a test of the Tobit specification against the above alternative suggested by Cragg. They challenged the assumption of the Tobit model by illustrating that older buildings are more likely to have fire incidence but are associated with lower economic damage, as new buildings usually have higher economic values. Cragg's two-part model relaxes the restriction of uniform effects in the participation and level decisions by separately implementing a probit for the participation decision and another standard regression for the level decision.

However, statistical analyses based on non-randomly selected samples can lead to erroneous conclusions. An example of such sample selection bias is self-selection. For example, "one observes wages for union members who found their non-union alternative less desirable" (Heckman 1979, pp153). Thus, "… wage or earnings functions estimated on selected samples, do not, in general, estimate population (i.e. random sample) wage functions".

Heckman (1979) shows that sample selection bias can be treated as an omitted variable issue. With the presence of sample selection bias, the error terms from the participation and level decisions are correlated (see Heckman 1979 for details). Heckman's sample-selection model assumes a bivariate normal distribution. The appeal of the sample selection model is that it also gives the unconditional mean and, thus, allows researchers to make inferences on the population, based on estimation from a non-randomly selected sample. The knowledge about such "potential" outcomes is valuable for policy implications. For example, health economic literature uses the sample selection model to study the consumption of alcohol and tobacco products, drawing policy implications to decrease consumption by the whole population.

As the maximum of the likelihood function for the sample selection model took a lot of time to compute in these days, Heckman suggests a two-step or limited information maximum likelihood (LIML) method. In the first stage, a probit regression estimates the probability of the dependent variable taking value of one. In the second stage, sample selection bias is corrected by incorporating a transformation of these predicted individual probabilities (called inverse Mill's ratio (IMR)) as an additional explanatory variable.

With the constant progress in computing power, full information maximum likelihood (FIML) estimation of Heckman's sample-selection model is later implemented and generally preferred, as FIML estimates are consistent and asymptotically efficient under the assumption of normality and homoscedasticity of the uncensored disturbances (i.e., the disturbances from the participation decision). See Puhani (2000) for a survey of Monte Carlo studies comparing the performance of the two methods.

In terms of statistical formula, the two-part model is a restricted form of Heckman's sample-selection model, where the correlation between two error terms is restricted to be zero. However, a simple significance test upon the correlation is not always a sufficient rule to choose between the two models. Instead, there are theoretical, practical and statistical grounds to choose between the two models. The next section, Section 2.3, describes the model selection procedures we follow.

Earlier studies using the above models retain to the analyses of single products. However, the consumption of some products is closely related, such as the consumption of tobacco and alcohol, and in our case, the consumption of whole and low-fat milk. Therefore, more recent development features a sample selection system or censored system for multiple-goods decisions, which allows correlation within and/or across participation decisions and level decisions among multiple goods. A number of censored-system estimation procedures have been proposed in the literature. These include maximum-likelihood estimators of Amemiya (1974), Wales and Woodland (1983), and Lee and Pitt (1986), and two-step estimators of Heien and Wessells (1990), Shonkwiler and Yen (1999), and Perali and Chavas (2000), as well as an extended full system approach of Stewart and Yen (2004) and Yen (2005).

Stewart and Yen’s (2004) multivariate sample-selection model (MSSM) is a generalization of the Tobit system (Amemiya 1974) and is also a multi-equation extension of Heckman’s sample-selection model, and its nested two-part model.

Depending on whether sample selection bias is present or not, two possible models are available for this study. They are bivariate two-part (B2P) model and multivariate sample-selection model (MSSM), which is the multivariate versions of the usual two-part model and Heckman’s sample-selection model, respectively.

To facilitate the presentation of models, we re-write Equations (4) and (5) in a system. Then each outcome variable $\mathrm{y_i}$ (expenditure on milk product i; whole milk if i=1, low-fat milk if i=2) is governed by a binary selection rule of whether to consume as follows (observation subscription omitted):

$$\begin{cases} \log(\mathrm{y_i}) = \mathbf{x}'\boldsymbol{\beta}_\mathbf{i} + \mathrm{v_i} & \text{if } \mathbf{z}'\boldsymbol{\alpha}_\mathbf{i} + \mathrm{u_i} > 0 \\ \mathrm{y_i} = 0 & \text{if } \mathbf{z}'\boldsymbol{\alpha}_\mathbf{i} + \mathrm{u_i} \leq 0, \quad \mathrm{i} = 1, 2 \end{cases} \tag{6}$$

where $\mathbf{z}$ and $\mathbf{x}$ are vectors affecting binary purchase decision and level decision, respectively; $\boldsymbol{\alpha}_\mathbf{i}$ and $\boldsymbol{\beta}_\mathbf{i}$ are vectors of parameters; $\boldsymbol{u}_\mathbf{i}$ and $\boldsymbol{v}_{\boldsymbol{i}}$ are random error in the participation and level equation, respectively.

To facilitate presentation of the log likelihood functions, define diagonal $\mathbf{S} = \mathrm{diag}(\sigma_1, \sigma_2)$ as standard deviation of $\boldsymbol{v}$. Let $\mathbf{R}_{\boldsymbol{uu}} = [\rho_{ij}^{uu}]$, $\mathbf{R}_{\mathbf{vu}} = [\rho_{ij}^{vu}]$, and $\mathbf{R}_{\boldsymbol{vv}} = [\rho_{ij}^{vv}]$ be 2 x 2 correlation matrices among elements of $\boldsymbol{u}$ and $\boldsymbol{u}$, $\boldsymbol{v}$ and $\boldsymbol{u}$, and $\boldsymbol{v}$ and $\boldsymbol{v}$, respectively.

2.2.1 Multivariate Sample-selection Model

The multivarite sample-selection model, proposed by Stewart and Yen (2004) and Yen (2005), extends Heckman's sample-selection model to a censored system involving multiple goods. MSSM assumes the concatenated error vector $[\boldsymbol{u}', \boldsymbol{v}']' \equiv [u_1, u_2, v_1, v_2]'$ is distributed as 4-variate normal with zero mean and covariance matrix:

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \tag{7}$$

where $\Sigma_{11} = \mathrm{E}(\boldsymbol{u}\boldsymbol{u}') = \mathbf{R}_{uu}$, $\Sigma_{21} = \Sigma_{12}' = \mathrm{E}(\boldsymbol{v}\boldsymbol{u}') = \mathbf{S}'\mathbf{R}_{vu}$, and $\Sigma_{22} = \mathrm{E}(\boldsymbol{v}\boldsymbol{v}') = \mathbf{S}'\mathbf{R}_{vv}\mathbf{S}$**.**

Define vectors $\boldsymbol{r} \equiv [r_1, r_2]' \equiv [z_1'\alpha_1, z_2'\alpha_2]'$ and$\boldsymbol{v} \equiv [\log(y_i) - \boldsymbol{x}'\boldsymbol{\beta_i}]$. Let $\phi(\boldsymbol{v})$ be the marginal probability density function (pdf) of $\boldsymbol{v} \sim N(0, \Sigma_{22})$ and $\phi(\boldsymbol{u}|\boldsymbol{v})$ be the conditional pdf of $\boldsymbol{u}|\boldsymbol{v} \sim N(\mu_{\boldsymbol{u}|\boldsymbol{v}}, \Sigma_{\boldsymbol{u}|\boldsymbol{v}})$, where $\mu_{\boldsymbol{u}|\boldsymbol{v}} = \Sigma_{12}\Sigma_{22}^{-1}\boldsymbol{v}$ and $\Sigma_{\boldsymbol{u}|\boldsymbol{v}} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$. Then, the likelihood contribution for the positive regime, where both dependent variables are positive, is given by:

$$L_1 = \phi(\boldsymbol{v}) \cdot \prod_{j=1}^{2} y_j^{-1} \cdot \int_{\boldsymbol{u}>-r}^{+\infty} \phi(\boldsymbol{u}|\boldsymbol{v}) d\boldsymbol{u} = \phi(\boldsymbol{v}) \cdot \prod_{j=1}^{2} y_j^{-1} \cdot \Phi_2(\boldsymbol{r} + \mu_{\boldsymbol{u}|\boldsymbol{v}}; \Sigma_{\boldsymbol{u}|\boldsymbol{v}}) \tag{8}$$

where $\prod_{j=1}^{2} y_j^{-1}$ is the Jacobian of the transformation from $[v_1, v_2]'$ to $[y_1, y_2]'$ and $\Phi_2(\boldsymbol{r} + \mu_{\boldsymbol{u}|\boldsymbol{v}}; \Sigma_{\boldsymbol{u}|\boldsymbol{v}})$ is the bivariate normal cumulative distribution function (cdf) with zero mean, covariance matrix $\Sigma_{\boldsymbol{u}|\boldsymbol{v}}$, and finite upper integration limits $\boldsymbol{r} + \mu_{\boldsymbol{u}|\boldsymbol{v}}$.

The second regime is one in which the values of both variables are zeros (when $\mathbf{z}'\boldsymbol{\alpha}_\mathbf{i} + \mathrm{u_i} \leq 0\,, \mathrm{i} = 1, 2$). The likelihood contribution is identical to that of an all-zero regime in the bivariate probit:

$$L_2 = \int_{-\infty}^{u \leq -r} \phi(\boldsymbol{u}, \Sigma_{11}) d\boldsymbol{u} = \Phi_2(-\boldsymbol{r}; \Sigma_{\mathbf{11}}) \tag{9}$$

where $\phi(\boldsymbol{u}, \Sigma_{11})$ is the marginal pdf of $\boldsymbol{u} \sim N(0, \boldsymbol{\Sigma_{11}})$. Specifically, $\phi(\boldsymbol{u}, \Sigma_{11}) = (2\pi)^{-1}|\Sigma_{11}|^{-1/2}\mathrm{e}^{-\frac{1}{2}\boldsymbol{u}'\boldsymbol{\Sigma_{11}}^{-1}\boldsymbol{u}}$.

For mixed regime, without loss of generality, denote u_i as the error term associated with the non-censored variable and u_j associated with the zero-valued variable. A mixed regime is characterized by:

$$\mathbf{z}'\boldsymbol{\alpha}_{\mathbf{i}} + \mathrm{u_i} > 0 \qquad \log(\mathrm{y_i}) = \boldsymbol{x}'\boldsymbol{\beta_i} + v_i \tag{10}$$

$$\mathbf{z}'\boldsymbol{\alpha}_{\mathbf{j}} + \mathrm{u_j} \leq 0 \qquad \mathrm{y_j} = 0.$$

Let $\tilde{v} \equiv v_i$, then $[\boldsymbol{u}', \tilde{v}]'$ is 3-variate normal with zero mean and covariance matrix $\widetilde{\boldsymbol{\Sigma}}$, where $\tilde{\Sigma}$ is a 3x3 sub-matrix containing the first three rows and columns of the error covariance matrix Σ in Equation (7). Partition $\tilde{\Sigma}$ at the third row and column such that $\widetilde{\boldsymbol{\Sigma}} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \widetilde{\boldsymbol{\Sigma}}_{12} \\ \widetilde{\boldsymbol{\Sigma}}_{21} & \widetilde{\boldsymbol{\Sigma}}_{22} \end{bmatrix}$.

Let $\phi(\tilde{v})$ be the marginal pdf of $\tilde{v} \sim N(0, \tilde{\Sigma}_{22})$ and $\phi(\boldsymbol{u}|\widetilde{\boldsymbol{v}})$ be the conditional pdf of $\boldsymbol{u}|\widetilde{\boldsymbol{v}} \sim N\ (\mu_{\boldsymbol{u}|\tilde{v}}, \Sigma_{\boldsymbol{u}|\tilde{v}})$, where $\mu_{\boldsymbol{u}|\tilde{v}} = \tilde{\Sigma}_{12}\tilde{\Sigma}_{22}^{-1}\widetilde{\boldsymbol{v}}$ and $\Sigma_{\boldsymbol{u}|\tilde{v}} = \Sigma_{11} - \tilde{\Sigma}_{12}\tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{21}$. Then the likelihood contribution for this regime is:

$$L_3 = y_i^{-1} \cdot \phi(\tilde{v}) \cdot \int_{u_i > -r_i}^{+\infty} \int_{-\infty}^{u_j \leq -r_j} \phi(u_1, u_2|\tilde{v}) du_2 du_1$$

$$= y_i^{-1} \cdot \phi(v_i) \cdot \Phi_2\big(\boldsymbol{D}(\boldsymbol{r} + \mu_{\boldsymbol{u}|\tilde{v}}; \boldsymbol{D}'\boldsymbol{\Sigma}_{\boldsymbol{u}|\tilde{v}}\boldsymbol{D}\big) \tag{11}$$

where $\boldsymbol{D} = diag(2d_1 - 1, 2d_2 - 1)$, $d_i = 1\ if\ \boldsymbol{z\alpha_i} + u_i > 0$. The sample likelihood function for the MSSM is the product of the likelihood contributions $\mathrm{L_1}$, $\mathrm{L_2}$, or $\mathrm{L_3}$ across observations, depending on the regime for each observation.

2.2.2 Bivariate Two-part Model

The bivariate two-part model (B2P) assumes no correlation between the participation and level equation, i.e. $\mathbf{R}_{vu} = \mathbf{0}$. It essentially constitutes a bivariate probit regression and a bivariate lognormal regression. However, consistent to the presentation of likelihood function for the MSSM model, we also partition the dataset into three regimes. The likelihood for the positive regime is:

$$L_1{}' = \phi(\boldsymbol{v}) \cdot \prod_{j=1}^{2} y_j^{-1} \cdot \int_{\boldsymbol{u}>-r}^{+\infty} \phi(\boldsymbol{u}) d\boldsymbol{u} = \phi(\boldsymbol{v}) \cdot \prod_{j=1}^{2} y_j^{-1} \cdot \Phi_2(\boldsymbol{r}; \Sigma_{\boldsymbol{22}}). \qquad (12)$$

The likelihood for negative regime, identical to that of an all-zero regime in the bivariate probit, is given by L_2 in Equation (9). For mixed regime, without loss of generality, denote u_i as the error term associated with the non-censored variable and u_j associated with the zero-valued variable. The likelihood function for the mixed regime is:

$$\mathrm{L}_3' = y_i^{-1} \cdot \phi(v_i) \cdot \int_{\boldsymbol{u_i}>r_i}^{+\infty} \int_{-\infty}^{\boldsymbol{u_j} \le r_j} \phi\big(\boldsymbol{u_i}, \boldsymbol{u_j}\big) \mathrm{d}\boldsymbol{u_j} \mathrm{d}\boldsymbol{u_i}$$

$$= y_i^{-1} \cdot \phi(v_i) \cdot \Phi_2(\boldsymbol{D}(\boldsymbol{r}); \boldsymbol{D}'^{\boldsymbol{\Sigma}_{22}}\boldsymbol{D}). \qquad (13)$$

The sample likelihood function for the bivariate two-part model is the product of $\mathrm{L}_1{}'$, L_2 or $\mathrm{L}_3{}'$ across observations, depending on the regimes of each observation. Taking logs of the sample probability function gives the log likelihood function. Estimation of the model by maximum likelihood estimation (MLE) is straightforward.

2.3 Model Selection

There is a well-established debate in health econometrics over the merits of Heckman's sample-selection models versus two-part models. We choose the B2P model after a comparison based on theoretical, practical and statistical grounds (Madden, 2008).

Firstly, the choice between the B2P and MSSM revolves around whether we wish to analyze the actual or the potential outcomes. For example, labor economists, who developed the Heckman's sample-selection model, are generally interested in the potential wage. Observations without positive wage outcomes do not imply that an individual worked for zero wages; instead they indicate that the potential wage (the wage that an individual could earn if she were to work) is unobserved. In the particular case of our study, what is the meaning of potential spending on milk? For those people with observed zero consumption of milk, is there a latent positive expected consumption which might have been incurred under certain circumstances? Milk consumption is traditionally a very important and common part of Polish diet. This nature of milk consumption in Poland leads us to believe that there is unlikely to be a latent positive expected consumption. Therefore, from a thereotical point of view, we choose the bivariate two-part model.

Secondly, there are also practical rules for estimation. The lack of exclusion restrictions in sample selection model usually leads to poor performance of its estimates. There is no exclusion restrictions if no variables that are in **x** are excluded from **z,** where **x** and **z** are the set of parameters in the level and participation equations, respectively. In these cases, the level equation is only identified through the nonlinearity of the IMRs.

However, collinearity problems are likely to prevail as IMR is an approximately linear function over a wide range of its argument **z** (Puhani 2000, pp57). The FIML estimator does not appear to depend on correlation between the IMR and the regressors, but Monte Carlo studies have shown that high collinearity also impairs the FIML estimator for the sample selection model (e.g. Leung and Yu 1996, pp213).

In our particular case of milk consumption, we believe the presence of children and elders might affect households' decision of whether to buy milk or not. But the level of expenditure is affected by the number of family members, as the total consumption quantity is directly determined by the size of a household. We further broke down the measure for family size into the numbers of children, adults, and elders, since these three groups generally have different nutritional needs. That is, these three variables enter the level equation only and thus serve as exclusion restrictions, if we were to use a sample selection model.

Leung and Yu (1996) have done a very detailed investigation into the performance of the sample selection model (LIML and FIML) and the alternative two-part model. The authors point out that (in the absence of exclusion restrictions) the degree of collinearity between the **x** regressors and the inverse Mill's ratio is the decisive criterion to judge the appropriateness of the sample selection model relative to the two-part model. In addition to the lack of exclusion restrictions, a small range of the argument of the inverse Mill's as well as high degree of censoring may also cause collinearity (e.g. Manning, Duan, and Rogers 1987; Zuehlke and Zeman 1991). Therefore, as a practical step, one should always check for collinearity even if there are exclusion restrictions.

Belsley (1991) provides a comprehensive list of diagnositics for analyzing collinearity in general. In our particular case, we are primarily concerned about the collinearity between the IMRs and the regressors **x**. So we adopt relatively simple diagnostics.We check the variance inflation factor (VIF) of the IMR in the level equation. The variance inflation factor $VIF = 1/(1 - R_i^2)$, whre R_i^2 is the determinant coefficients from the auxilary regression of x_i regressed on the remaining explnanatory variables. A sufficient condition of the presence of collinearlity for a particular regessor is a high VIF. What precisely defines “high” VIF is open to quesiton, but Belsley (1991) suggests a value of 30.

In the particular case of this study, the VIF of IMR in the level equation for whole milk is 23.82, while that in the equaiton for low-fat milk is as high as 147.45. It seems that there is no severe collinearity problem for whole milk consumption, but the IMR is highly correlated with the regressors **x** for low-fat milk consumption.

The practical ground rules out the sample selecton model for low-fat milk consumption, but not for whole-milk consumption. But it might be worthy to estimate the sample selection model and formally test against the two-part model.

Lastly, there may be statistical tests to discriminate between the two models. One commonly used criterion in Monte Carlo studies is the mean square error (MSE) of the parameter of interest (e.g. Leung and Yu 1996). The MSE is the variance plus the square of the bias. Knowledge of the true parameter is needed to compute the bias, but in empirical application the true parameter values are unknown. In this situation Dow and Norton (2003) recommend the test proposed by Toro-Vizcarrondo and Wallace (1968) for application with multicollinearity, which they call an empirical MSE test. This involves calculating the emprical MSE of both estimators under the assumption that one model, e.g. the two-part model, is consistent and correct. The MSE for the two-part model will then involve only the variance component, while that for the selection model will involve its variance and its "bias" relative to the assumed true model.

Dow and Norton (2008) show through Monte Carlo simulations that the empirical MSE test upon parameter estimates yields the same model choice as does the true MSE criterion. This thesis calculates and compares the empirical MSE on the marginal effect of each explanatory variable, a primary interest of this thesis.

2.4 Marginal Effects

Economically meaningful measure, marginal effects, are calculated based on conditional means for the joint distribution. The probability of purchase is given by:

$$\mathrm{Pr}(\mathrm{y_i} > 0) = \Phi(\mathbf{z}'\boldsymbol{\alpha}_\mathbf{i}). \tag{14}$$

Elasticity for continuous explanatory variable is defined as the change in probability of purchase, corresponding to a one-unit change in $\mathrm{z_j}$. The marginal effects for indicator explanantory variables are the discrete change in purchase probabilties obtained in Equation (14) when the explanatory variable takes value of one versus zero:

$$\mathrm{m_i^{Prob}} = \begin{cases} \frac{\mathrm{d\ Pr(y_i>0)}}{\mathrm{dz_j}} = \phi(\mathbf{z}'\boldsymbol{\alpha}_\mathbf{i}) \cdot \alpha_{\mathrm{ij}}, & if\ z_j\ continuous \\ \Phi\big(\mathbf{z}'\boldsymbol{\alpha}_\mathbf{i}\big|\mathbf{z_j} = 1\big) - \Phi\big(\mathbf{z}'\boldsymbol{\alpha}_\mathbf{i}\big|\mathbf{z_j} = 0\big), & if\ z_j\ binary \end{cases} \tag{15}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the pdf and cdf of the standard normal distribution, respectively.

The conditional mean of expenditure y_i is (Rosiniski and Yen, 2004):

$$E(y_i|y_i>0)=\begin{cases}\exp\left(\boldsymbol{x}'\boldsymbol{\beta}_{\boldsymbol{i}}+\frac{\sigma_i^2}{2}\right)\cdot\Phi(\boldsymbol{z}'\boldsymbol{\alpha}_{\boldsymbol{i}}+\rho_{ii}^{uv}\sigma_i^2)/\Phi(\boldsymbol{z}'\boldsymbol{\alpha}_{\boldsymbol{i}}),\ MSSM\\ \exp\left(\boldsymbol{x}'\boldsymbol{\beta}_{\boldsymbol{i}}+\frac{\sigma_i^2}{2}\right),\ B2P.\end{cases}\quad(16)$$

Multiplying Equations (14) and (16) gets the unconditional mean of y_i:

$$E(y_i)=\begin{cases}\exp\left(\boldsymbol{x}'\boldsymbol{\beta}_{\boldsymbol{i}}+\frac{\sigma_i^2}{2}\right)\cdot\Phi(\boldsymbol{z}'\boldsymbol{\alpha}_{\boldsymbol{i}}+\rho_{ii}^{vu}\sigma_i^2), MSSM\\ \exp\left(\boldsymbol{x}'\boldsymbol{\beta}_{\boldsymbol{i}}+\frac{\sigma_i^2}{2}\right)\cdot\Phi(\boldsymbol{z}'\boldsymbol{\alpha}_{\boldsymbol{i}}), B2P.\end{cases}\quad(17)$$

Let's consider a variable that enters the level equation as well as the participation equation. In this case, when deriving the semi-elasticity of conditional expected value of y_i with respect to $\mathrm{x_j}$, we have to consider that vector **z** also contains $\mathrm{x_j}$.

In the MSSM model, semi-elasticity (discrete change) of the conditional mean is obtained by differentiating (differencing) Equation (15) with respect to variable x_j:

$$\mathrm{m_i^c}=\begin{cases}\frac{\mathrm{d}\ \ln E(y_i|y_i>0)}{\mathrm{dx_j}}=\beta_{\mathrm{ij}}+[\lambda(\boldsymbol{z}'\boldsymbol{\alpha}_{\boldsymbol{i}}+\rho_{ii}^{vu}\sigma_i)-\lambda(\boldsymbol{z}'\boldsymbol{\alpha}_{\boldsymbol{i}})]\alpha_{\mathrm{ij}}\ ,\ if\ x_j\ continuous\\ \Delta\ \ln E(y_i|y_i>0)=\beta_{\mathrm{ij}}+\Delta[\lambda(\boldsymbol{z}'\boldsymbol{\alpha}_{\boldsymbol{i}}+\rho_{ii}^{vu}\sigma_i)-\lambda(\boldsymbol{z}'\boldsymbol{\alpha}_{\boldsymbol{i}})]\ ,\ if\ x_j\ binary\end{cases}\quad(18)$$

where α_{ij} and β_{ij} are the parameters of x_j in the participation equation and level equation for milk product i, respectively; $\Delta[\cdot]$indicates the difference of its argument when x_j takes value of one versus zero. And the inverse Mill's ratio $\lambda(\boldsymbol{z}'\boldsymbol{\alpha}_{\boldsymbol{i}})\equiv\frac{\phi(\boldsymbol{z}'\boldsymbol{\alpha}_i)}{\Phi(\boldsymbol{z}'\boldsymbol{\alpha}_i)}$. In the B2P model, $\rho_{ii}^{vu}=0$, and, thus, the semi-elasticity of the conditional expected expenditure reduces to parameter β_{ij}, with respect to either continuous or binary variable.

For the MSSM model, semi-elasticity (discrete change) of the unconditional mean with respect to x_j that enters both equations is obtained by differentiating (differencing) Equation (16):

$$\mathrm{m_i^u} = \begin{cases} \frac{\mathrm{d}\ \mathrm{ln}E(y_i)}{\mathrm{dx_j}} = \mathrm{\beta_{ij}} + \ \lambda(\boldsymbol{z}'\boldsymbol{\alpha_i} + \rho_{ii}^{vu}\sigma_i)\alpha_{ij},\ if\ x_j\ is\ continuous \\ \Delta\ \mathrm{ln}E(y_i) = \mathrm{\beta_{ij}} + \Delta[\lambda(\boldsymbol{z}'\boldsymbol{\alpha_i} + \rho_{ii}^{vu}\sigma_i)],\ if\ x_j\ binary. \end{cases} \qquad (19)$$

For variables that enter the level equation only, the marginal effects for conditional and unconditional mean under both models are its parameter $\mathrm{\beta_{ij}}$ only.

In this thesis, individual elasticity or discrete change is averaged over the whole sample to obtain the average marginal effect. Asymptotic standard errors for the average marginal effect estimates are obtained using the delta method (Spanos, 1999).

CHAPTER 3

DATA DESCRIPTION

The data are from the Polish household panel of about 20,000 households annually surveyed by Poland's Main Statistical Office (GUS). Despite the attempted panel structure of the survey, fewer than 36% of the households were observed for more than one year. We use a pooled cross-sectional sample of 108,064 observations with non-missing values for the period of 2004 to 2008. The aggregate numbers of outmigration by regions are also obtained from GUS.

Table 1 presents variable definitions and sample statistics. The percent of households with non-zero expenditure on whole and low-fat milk is 61.2% and 62.4%, respectively. And 29.5% of the total 108,064 observations make nonzero purchase on both milk products at the same time. The average expenditures on whole and low-fat milk over the whole sample are 12.02 Polish Zloty (PLN) and 9.01 PLN respectively.

Table 1. Variable definitions and Sample Statistics

Variable	Description/Unit	Mean	SD
	Food Expenditures / Dependent variables		
Buy1	1, if household buys whole milk, 0 otherwise	0.612	0.487
Buy2	1, if household buys low fat milk, 0 otherwise	0.624	0.484
Wmilk	Expenditure on whole milk in the month preceding survey, in Polish Zloty (PLN)	12.016	9.830
Lmilk	Expenditure on low fat milk in the month preceding survey (PLN)	9.007	3.555
	Demographic, Socio-Economic Factors/Explanatory variables		
Village	1, if a household residents in village, 0 otherwise	0.362	0.481
Income	Household income in the month preceding survey (1000 PLN)	2.306	1.473
Male	1, if the household head is male, 0 otherwise	0.595	0.491
Married	1, if the household head is married, 0 otherwise	0.678	0.467
Educ	1, if the household head has secondary or higher education, 0 otherwise	0.402	0.490
Age	Household head's age, in years	50.888	15.288
Employ	1 if household head is permanently employed or contract employee, 0 otherwise	0.421	0.494
Child	Number of children (under 18 years old)	0.639	0.985
Adult	Number of adults 60 or under 60 years old	1.823	1.198
Elder	Number of elders above 60 years old	0.451	0.695
OutD	Net migration domestically to other regions in Poland, in 1000	-1.252	5.598
OutF	Net migration internationally to other countries, in 1000	1.333	2.053
Time	1-5 corresponding to year 2004 to 2008	2.762	1.473

N=108,064

Rural residents account for 36.2% of all observed households. Household income in the month preceding survey averages 2,306 PLN. Male members are heads in 59.8% of households; and 67.8% of household heads are married. The proportion of household heads that have a secondary or higher education is 40.2%.

The average household head's age is 50.9 years. In terms of employment stability, 42.1% household heads are permanently employed or contract employees. The average number of children (age 0-18), adults (age 19-60), and elders (above 60) is 0.64, 1.82 and 0.45 per household, respectively.

Two variables are reported as measure of depopulation. First, net domestic migration (*OutD*) measures the net outflow of population from a region to other regions within Poland. Second, net international migration (*OutF*) measures the net outflow of population from a region to other countries. The 16 Polish administrative regions on average experience a net migration inflow about 1,252 persons, while net international outmigration averages 1,333 persons.

About 30% households are observed in 2004 and about 17%~18% households are observed in each year from 2005 to 2008. A new continuous variable *Time* (values 1 to 5) is created to capture time trend in milk expenditures.

CHAPTER 4

RESULTS

This chapter reports the estimation results. Section 4.1 compares the MSSM and B2P model by the use of empirical MSE tests. Section 4.2 reports the marginal effects under the chosen model, while Section 4.3 presents goodness of fit measures.

4.1 Estimates and Comparison of MSSM and B2P model

Maximum likelihood parameter estimates for MSSM and B2P model are obtained, respectively (Tables A1 and A2 in Appendix). Each explanatory variable's marginal effect on purchase probability and expenditure level is estimated. The corresponding asymptotic standard errors are obtained by the delta method. Furthermore, in order to choose the appropriate model, empirical MSE of above marginal effects under both models are computed and compared.

Table 2 presents the empirical MSE associated with the marginal effects under MSSM and B2P model, respectively, under the hypothesis that the B2P model is the "true" model. The empirical MSE are then calculated as follows (Dow and Norton 2008, 15):

$$\mathrm{EMSE(B2P)} = \mathrm{var(B2P)} + (\mathrm{m}^{\mathrm{B2P}} - \mathrm{m}^{\mathrm{B2P}})^2$$

$$\mathrm{EMSE(MSSM)} = \mathrm{var(MSSM)} + (\mathrm{m}^{\mathrm{MSSM}} - \mathrm{m}^{\mathrm{B2P}})^2$$

Where $\mathrm{m}^{\mathrm{B2P}}$ and $\mathrm{m}^{\mathrm{MSSM}}$ are the estimated marginal effects under B2P and MSSM, respectively.

Table 2. Empirical MSE of Marginal Effects under H_0: B2P Model Is True

Empirical MSE of marginal effect on purchase probability						
	Whole milk			Low-fat milk		
	MSE	MSE		MSE	MSE	
Variable	MSSM	B2P	Choice	MSSM	B2P	Choice
Income	0.142	0.066	B2P	0.075	0.137	MSSM
Educ	2.387	0.359	B2P	1.040	0.003	B2P
Age	2.513	0.000	B2P	3.045	0.004	B2P
Male	2.440	0.000	B2P	6.647	1.002	B2P
Married	0.017	3.531	MSSM	0.879	0.949	MSSM
Employ	7.486	0.258	B2P	0.968	1.532	MSSM
Village	3.059	18.293	MSSM	64.592	30.441	B2P
Dchild	0.528	4.223	MSSM	4.575	0.352	B2P
Delder	4.890	0.978	B2P	14.481	0.492	B2P
OutD	13.632	0.045	B2P	10.809	0.037	B2P
OutF	12.069	0.102	B2P	15.575	0.023	B2P
Time	13.628	0.178	B2P	21.188	0.238	B2P
Empirical MSE of marginal effect on conditional expected expenditure						
	Whole milk			Low-fat milk		
	MSE	MSE		MSE	MSE	
Variable	MSSM	B2P	Choice	MSSM	B2P	Choice
Income	0.772	0.143	B2P	0.048	0.553	MSSM
Educ	3.364	1.141	B2P	3.238	0.351	B2P
Age	1.717	0.001	B2P	1.327	0.014	B2P
Male	0.490	4.452	MSSM	0.416	0.440	MSSM
Married	0.498	2.264	MSSM	0.594	9.740	MSSM
Employ	46.183	21.055	B2P	44.827	1.598	B2P
Village	163.697	236.477	MSSM	149.555	8.318	B2P
Child	9.329	38.686	MSSM	5.566	29.542	MSSM
Adult	4.123	28.956	MSSM	0.250	11.025	MSSM
Elder	22.063	71.928	MSSM	4.615	31.010	MSSM
OutD	14.939	0.001	B2P	12.210	0.118	B2P
OutF	21.324	0.347	B2P	26.333	1.293	B2P
Time	13.443	0.445	B2P	18.255	0.002	B2P

Note: MSE are multiplied by 1000

For each marginal effect, a model with lower MSE is chosen. For purchase probability, the B2P estimate is chosen 18 times out of 24 cases. And for the conditional expected expenditure, the B2P estimate is chosen 14 times out of 26 cases. This result of empirical MSE test is consistent with our choice based on theoretical and practical reasons.

4.2 Results under the Final Model (B2P)

Since B2P model is the ultimately chosen model, the following sections describe the results of B2P model. Table 3 reports each explanatory variable's effect on the probability of purchase and the level of milk expenditure, as well as the estimates of correlation coefficients.

Consistent with expectation, the decision to buy the two milk products under study as well as the expenditure amounts are negatively correlated, as reflected by the negative correlation coefficient estimates of -0.578 and -0.016, respectively. These estimates are statistically significant with *p*-value less than 0.001.

The correlation coefficient between the level equations is small (-0.016) because we pooled single-product buyers into the correlation estimation. As described in Chapter 3, the percentage of households that have nonzero expenditure on whole and low-fat milk is 61.2% and 62.4%, respectively. About 47% of these milk buyers make nonzero expenditure on both milk products; but the rest of them are single-product buyers. For those "double" buyers, the correlation coefficient is as high as -0.141 (p<0.0001), indicating a clear substitution between whole and low-fat milk. However, we are not only interested in those "double" buyers, but also in the whole sample. Therefore our B2P model allows the nonzero whole milk expenditure to be correlated with the nonzero low-fat milk expenditure, regardless of "single" or "double" buyers.

Once counting for those “single” buyers, the correlation coefficient is reduced to -0.016, as we have a substantially high percentage of single buyers.

The signs of all variables are not the same across the participation and level equations. For example, higher international outmigration (*OutF*) increases the probability of purchase whole milk but is negatively correlated with the conditional mean of expenditure of whole milk. This result implies that different decision rules are applied when households decide whether to buy milk and if they buy, how much to spend. Such decision-making process mandates the use of two different equations, as recognized in our modeling scheme.

For most demographic variables, the signs of marginal effects are consistent with expectations and previous results reported in literature, as described in the remainder of this section. In the following report of estimation results, the effects of each explanatory variable are compared across the purchase and amount decisions.

In the binary decision to buy milk, income positively influences the probabilities of buying whole and low-fat milk. The purchase probabilities increase by 0.81% and 1.17% for whole and low-fat milk, respectively, when income increases by 1000PLN. Holding other variables constant, more affluent households spend more on low-fat milk, but less on whole milk. Specifically, if household income increases by 1000PLN, the expenditure on whole milk declines by 1.19%, while that on low-fat milk increases by 2.35%. Low-fat milk has the same nutritional benefits as whole milk but contains less saturated fat and is generally considered healthier. In other words, households with higher income make relatively healthier milk consumption decisions.

Table 3. Marginal Effects under Bivariate Two-part (B2P) Model

Marginal effect on purchase probability							
	Whole milk				Low-fat milk		
Variable	Estimate	(se*100)			Estimate	(se*100)	
Income	0.81%	(0.175)	***		1.17%	(0.252)	***
Educ	-1.89%	(0.319)	***		-0.16%	(0.333)	
Married	5.94%	(0.375)	***		3.08%	(0.373)	***
Age	-0.05%	(2.923)			0.19%	(4.003)	
Male	0.06%	(0.342)			-3.17%	(0.350)	***
Employ	-1.60%	(0.343)	***		3.91%	(0.351)	***
Village	13.53%	(0.306)	***		-17.45%	(0.345)	***
Dchild	6.50%	(0.330)	***		1.88%	(0.343)	***
Delder	3.13%	(0.425)	***		-2.22%	(0.418)	***
OutD	-0.67%	(0.140)	***		0.61%	(0.031)	***
OutF	1.01%	(0.106)	***		-0.48%	(0.145)	***
Time	1.33%	(0.197)	***		-1.54%	(0.229)	***
Marginal effect on conditional expected expenditure							
	Whole milk				Low-fat milk		
Variable	Estimate	(se*100)			Estimate	(se*100)	
Income	-1.19%	(0.310)	***		2.35%	(0.316)	***
Educ	-3.38%	(0.862)	***		1.87%	(0.826)	**
Married	4.76%	(1.043)	***		9.87%	(1.040)	***
Age	-0.10%	(0.034)	***		0.38%	(0.035)	***
Male	6.67%	(0.892)	***		2.10%	(0.873)	***
Employ	-14.51%	(0.906)	***		-4.00%	(0.901)	***
Village	48.63%	(0.865)	***		9.12%	(0.887)	***
Child	19.67%	(0.427)	***		17.19%	(0.478)	***
Adult	17.02%	(0.447)	***		10.50%	(0.483)	***
Elder	26.82%	(0.800)	***		17.61%	(0.901)	***
OutD	-0.09%	(0.069)	*		1.09%	(0.077)	***
OutF	-1.86%	(0.198)	***		-3.60%	(0.197)	***
Time	2.11%	(0.262)	***		0.14%	(0.265)	
Correlation coefficients and variance estimates							
Rho.buy1.buy2	-0.578	(0.583)	***	Rho.y1.y2	-0.016	(0.556)	***
Sigma.y1	0.964	(0.235)	***	Sigma.y2	0.965	(0.222)	***

Note: *** *P*<0.01; ** *P*<0.05; * *P*<0.10

Household heads with higher education are less likely to buy whole milk. They also spend less on this product and more on low-fat milk, compared to their counterparts with lower education attainment level. People with higher education are more likely to recognize the relative superiority of low-fat milk and, therefore, spend more on it.

Married household heads are more likely to buy, and spend more as well, on both milk products. The age of household heads does not affect the probabilities of buying whole or low-fat milk. But older household heads are associated with higher expenditure on low-fat milk and less on whole milk. This is consistent with the finding in literature that there is an association between increasing age and adherence to a low-fat diet (e.g. Nigg et al. 1999).

The variable male household head (*Male*) is insignificant regarding the purchase decision of whole milk. But male household heads are 3.17% less likely to buy low-fat milk than female household heads. This is consistent with findings reported in literature that females are generally more concerned about diet healthiness, especially with regard to fat intake (Wardle et al. 2004). However, once they have decided to buy milk, male household heads on average spend 6.67% and 2.10% more on whole and low-fat milk, respectively, than their female counterparts.

Households with relatively stable employment are less likely to buy whole milk, but more likely to buy low-fat milk. Moreover, employment stability is related to lower milk expenditures, possibly because these household heads have different diet composition. They are more likely to skip breakfast, or because these households eat away-from-home more often.

Households residing in villages are 13.53% more likely to buy whole milk and have a 17.45% lower probability of buying low-fat milk, reflecting relatively inferior milk choice in rural areas. Conditional on purchase, however, rural residents spend significantly more on both milk products. Rural households spend 48.63% and 9.12% more on whole and low-fat milk, respectively, than their urban counterparts. The particularly strong preference for whole milk may call for additional research.

The presence of children is estimated to increase both the probabilities of buying whole and low-fat milk. While the presence of elder people increases the probability of buying whole milk, it decreases the probability of buying low-fat milk. Note this is not inconsistent with the finding that older household heads spend more on low-fat milk, because an elder family member is not necessarily the household head. It is plausible that the elders are parents of the household head and consume whole milk by habit because low-fat milk was less accessible when they were growing up in Poland.

Larger numbers of family members (children, adults, and elders, respectively) are also associated with higher expenditure on both milk products. Although the effects of these three measures are of the same sign, their values vary. Therefore, the decomposition of family size into three categories provides insights about the different weight each factor carries in a household's decision to buy whole and/or low-fat milk.

Worker outmigration shows an interesting effect. Households residing in regions with higher domestic outmigration are less likely to buy whole milk and more likely to buy low-fat milk. They also spend less on whole milk and more on low-fat milk. Domestic out-migrants are typically young, well-educated and they are less likely to consume whole milk, more likely to consume low-fat milk. Their choices are possibly communicated back to households where they have come from, which leads to the change in preferences as revealed by our empirical analyses.

In contrast, higher international outmigration is associated with higher probability of whole milk purchase, but lower probability of buying low-fat milk. Conditional on purchase, households in regions with higher international outmigration spend less on both milk products. Migration that happens in Poland is mostly job migration in search of higher wages. Foreign out-migrants tend to be less educated, young and middle-aged job seekers. If whole milk has been established as their milk choice, the households they left continue to prefer whole milk, because they are likely households of less educated and lower. Any remittances from abroad that add to household income in Poland support the consumption pattern in terms of milk choices, but they do not change the choice of milk type.

As above discussed, the possible underlying mechanism for the different effects of outmigration can be very different. For example, the unbalanced regional economic development or foreign remittances reflected in the household income might significantly shift consumption patterns. Foreign cultural and life-style exposure coupled with outmigration may also contribute to the changes in consumption behavior.

Since regions with high outmigration tend to be economically less developed, the supported development of dairy sector (at least milk products) in these areas may provide the potential of additional employment and economic development.

Lastly, the expenditure on whole milk has been increasing over the study period. This is a promising trend, contrasting with the declining fluid milk consumption in developed countries. However, the probability of buying whole milk is increasing over the years, while that of buying low-fat milk is decreasing. In addition, there is no statistically significant increase over time in the expenditure on low-fat milk, either. This is a trend worth of notice for public health policies, as fluid milk (except for skim milk) also contains saturated fatty acids, which if consumed in large quantities over extended period of time, are associated with declining circulatory health in humans.

4.3 Goodness of Fit Measures and Residual Analyses

We check the goodness of fit for the B2P model by inspecting the two parts separately. Table 4 reports the goodness of fit measures for the participation decisions.

Table 4. Goodness of Fit Measures for Bivariate Probit Regression

Log likelihood	-131936.954	
Log Likelihood with intercept only	-136741.000	
McFadden R^2 $(1 - LL_{Full}/LL_{Intercept})$	0.0351	
McFadden's adjusted R^2$(1 - (LL_{Full} - K)/LL_{Intercept})$	0.0350	
LR test for the overall significance (df=24)	9608.092	p<0.0001

	Whole milk	Low-fat milk	Overall
Percent concordant	62.250%	63.898%	63.074%

As typical in cross sectional analysis, McFadden's (1980) R-squared is quite low (0.0351). However, the likelihood ratio test suggests joint significance of all explanatory variables. The overall proportion of correctly predicted is 63.1%, while that for decisions to buy whole and low-fat milk are 62.3% and 63.9%, respectively. The overall significance of the model indicates that the model provides satisfactory explanatory power and fits the data reasonably well.

Detailed comparisons between the observed and expected probabilities are given in Tables 5 and 6, for whole and low-fat milk equations, respectively. The sample is divided into ten groups based on the estimated probabilities, a grouping rule proposed by Hosmer and Lemeshow (1980) and Lemeshow and Hosmer (1982) when there is continuous explanatory variable. As shown in the bolded columns in both tables, for each group, the predicted probabilities are very close to the observed probabilities.

Table 5. Observed versus Expected Probabilities and Frequencies, Whole Milk

Decile	Cut point	N obs	Observed probability	Predicted probability
1	48.684%	10806	44.97%	45.64%
2	52.255%	10807	50.52%	50.54%
3	55.187%	10807	53.75%	53.73%
4	57.992%	10806	56.67%	56.63%
5	60.524%	10807	60.20%	59.26%
6	63.708%	10804	63.01%	62.05%
7	66.942%	10808	66.01%	65.31%
8	70.770%	10812	68.28%	68.87%
9	74.599%	10800	71.01%	72.62%
10	87.393%	10807	77.27%	77.79%
Total		108064	61.17%	61.24%

Table 6. Observed versus Expected Probabilities and Frequencies, Low-fat Milk

Decile	Cut point	N obs	Observed probability	Predicted probability
1	47.767%	10807	42.78%	43.85%
2	51.663%	10805	49.88%	49.83%
3	55.288%	10810	54.04%	53.43%
4	60.758%	10804	58.88%	57.78%
5	65.313%	10806	63.11%	63.34%
6	67.833%	10804	66.23%	66.64%
7	69.890%	10806	68.99%	68.89%
8	71.711%	10808	71.75%	70.80%
9	73.870%	10807	73.39%	72.74%
10	85.602%	10808	74.77%	75.84%
Total	–	108064	62.38%	62.31%

For the bivariate lognormal regression, the common goodness of fit measure, R^2, are reported for the expenditure on whole and low-fat milk, respectively (Table 7). The model explains 99.87% and 99.57% of the variations in the expenditure on whole and low-fat milk, respectively. The LR test has a *p*-value lower than 0.0001, and thus concludes good overall fit.

Table 7. Goodness of Fit Measures for Bivariate Lognormal Regression

	Whole milk	Low-fat milk	Overall
n	66,102	67,411	101,684
R^2	0.9987	0.9957	
Log likelihood	-560,965.9664		
Log likelihood with intercept only	-619,998.0294		
LR test for overall significance (df=26)	118,064.1260	*p*-value< 0.0001	

Additional residual analyses are performed to check the assumption of normality and homoscedasticity. Figure 1 presents QQ plots and histogram of the residuals for both milk products. Although the model residuals deviants from normal distribution, they are generally bell-shaped, indicating that the assumption of normality is reasonably made.

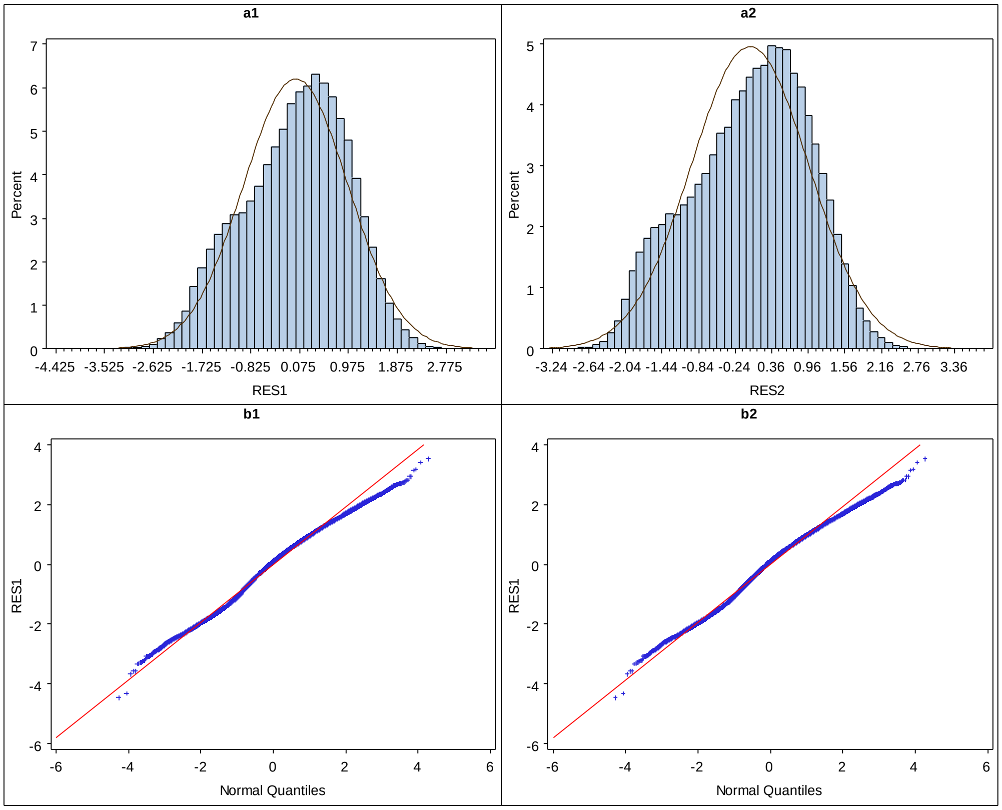

Figure 1. Normality Test on the Residuals of the Bivariate Lognormal Regression

Lastly, residual plots against each of the 13 explanatory variables rule out the presence of heteroscedasticity. See Figures A1 to A4 in the Appendix for details. In summary, we compute goodness of fit measures and perform residual analyses to check the appropriateness of our model. The results provide satisfactory evidence of good fit.

CHAPTER 5

CONCLUSION AND DISCUSSION

Dairy products, mainly milk products, are a preeminent food category in the retail sector of Poland. Milk is also a major source of essential nutrients and an important part of Polish diet. Given the importance of dairy sector and milk's essential role in Polish diet, we investigate the effects of demographic and socio-economic factors on the decisions to consume whole and low-fat milk. The variables scrutinized include household head's gender, age, education level, marital status and employment stability, household location, monthly income, and the numbers of children (age 0-18), adults (age19-60) and elders (age >60). Two additional variables, *OutD* and *OutF*, measure the net domestic and international outmigration in a region. And lastly, a time trend is included.

A system approach is used to allow the possible correlation between the consumption of whole and low-fat milk. Out of the two possible models, bivariate two-part model is chosen over multivariate sample -selection model based on theoretical, practical and statistical grounds.

The findings from model estimation are consistent with expectations and literature. In the decision whether to buy whole and/or low-fat milk, households with higher income, larger size and married heads have higher probabilities buying both milk products. Higher education, older household heads, and more stable employment are associated with relatively healthier milk choice.

However, rural residents usually appear to make less healthy fluid milk choice, as they are more likely to buy whole milk. Male household heads are also less likely to buy low-fat milk.

Worker outmigration has mixed effects. With regard to the probability of purchase, domestic outmigration is associated with healthier choice, while higher international outmigration is related to higher probability of buying whole milk and lower probability of buying low-fat milk. The possible underlying reason includes economic factors, different foreign cultural and life-style exposure, and a different migration pattern. This could be a topic for future research to study the dietary welfare of the population in those regions.

Our empirical analyses also reveal the trend of milk consumption in Poland over the years, providing insights for public health policies. The upward trend in whole milk consumption is somewhat optimistic, in contrast to the declining fluid milk consumption in developed world. Low-fat milk is considered healthier because of its lower level of saturated fat. However, the purchase probabilities of low-fat milk among Polish households are declining over the years. This is a trend worthy note and requires examination and possible intervention.

In summary, the bivariate two-part model enables a close scrutiny of the demographic and socio-economic factors affecting household whole and low-fat milk consumption in Poland. The resulted findings revealed the direction of each variable’s effect as well as its magnitude. The findings are important to learn about which factors are associated with healthy (unhealthy) milk choice and are informative for the formulation of economic and public health policies. Given the importance of dairy sector and the nutritional value of milk, this study provides a notion about potential of dairy sector (milk products) for local employment and economic development.

REFERENCES

Amemiya, T. 1985. Advanced Econometrics. Cambridge: Harvard University Press.

Belsley, D. 1991. Conditional Diagnostics: Collinearity and Weak Data in Regression. John Wiley and Sons, New York.

Bettman, J.R. 1979. “Memory Factors in Consumer Choice: A Review.” *Journal of Marketing* 43:37–53.

Cragg, J.G. 1971. “Some Statistical Models for Limited Dependent Variables with Applications to the Demand for Durable Goods.” *Econometrica* 39(1971):829–844.

Dow, W., and E. Norton. 2003. “Choosing Between and Interpreting the Heckit and Two-part Models for Corner Solutions.” *Health Services and Outcomes Research Methodology* 4 (1): 5–18.

Gensch, D.H. 1987. “A Two Stage Disaggregate Attribute Choice Model.” *Marketing Science* 6 (Summer 1987):223–231.

GUS (Main Statistical Office) 2013. Budzety gospodarstw domowych 2012. Warszawa.

____. (Main Statistical Office) 2010. Budzety gospodarstw domowych 2009. Warszawa.

Heckman J.J. 1976. “The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Department Variables and a Simple Estimator for Such Models.” *Annals of Economic Social Measurement* 5 (4):475–492.

______. 1979. “Sample Selection Bias as a Specification Error.” *Econometrica* 47 (1) (Jan., 1979):153–161.

Heien, D. and C.R. Wessells.1990. “Demand Systems Estimation with Microdata: a Censored Regression Approach.” *Journal of Business and Economic Statistics* 8(3):365-371.

Hosmer, D.W. and S. Lemeshow. 1980. “Goodness of Fit Test for the Multiple Logistic Regression Model.” *Communications in Statistics – Theory and Methods* 9(10):1043–1069.

Klepacka A. M., T. S. Sobczynski, W. J. Florkowski, and C. Revoredo-Giha. 2013. Effects of the Common Agricultural Policy a Non-Family Farm Employment in Primary and Secondary Agricultural Areas. Paper presented at the joint AAEA/CAEA annual meeting, Washington, D. C., August 4 – 6, 2013.

Lee, L-F., and M.M. Pitt. 1986. “Microeconometric Demand Systems with Binding Nonnegativity Constraints: The Dual Approach.” *Econometrica* 54:1237–42.

Lemeshow, S. and D.W. Hosmer. 1982. “A Review of Goodness of Fit Statistics for Use in the Development of Logistic Regression Models.” *American Journal of Epidemiology* 115(1):92 – 106.

Leung, S.F. and S. Yu. 1996. “On the Choice between Sample Selection and Two-part Models.” *Journal of Econometrics* 72:197–229.

Lin T.F. and P. Schmidt. 1984. “A Test of the Tobit Specification against an Alternative Suggested by Cragg.” *The Review of Economics and Statistics* 66(1):174–177.

Madden, D. 2008. “Sample Selection versus Two-part Models Revisited: The Case of Female Smoking and Drinking.” *Journal of Health Economics* 27 (2008):300–307.

Manning, W.G., N. Duan, and W.H. Rogers. 1987. “Monte Carlo Evidence on the Choice Between Sample Selection and Two-Part Models.” *Journals of Econometrics* 35:59–82.

McFadden, D. 1980. “Econometric Models for Probabilistic Choice among Products.” *Journal of Business* 53(3) Part 2: Interfaces between Marketing and Economics: 13–29.

Nigg, C.R., P.M. Burbank, C. Padula, R. Dufresnem J.S. Rossi, W.F. Velicer, R.G. Laforge and J.O. Prochaska. 1999. “Stages of Change across Ten Health Risk Behaviors for Older Adults.” *The Gerontologist* 39(4):473-482.

Perali, F., and J.P. Chavas. 2000. “Estimation of Censored Demand Equations from Large Cross-Section Data.” *American Journal of Agricultural Economics* 82:1022–1037.

Puhani, P.A. 2000. “The Heckman Correction for Sample Selection and Its Criteque.” *Journal of Economic Surveys* 14(1):53-68.

Rosinski, J., and S.T. Yen. 2004. “A Note on the Conditional Moments of Limited Dependent Variable Models with a Transformed Dependent Variable.” Unpublished, Dept. Agr. Econ., The University of Tennessee, Knoxville, 2004.

Shocker A.D., M. Ben-Akiva, B. Boccara, and P. Nedungadi. 1991. “Consideration Set Influences on Consumer decision-Making and Choice: Issues, Models and Suggestions.” *Marketing Letters*, 2(Aug., 1991):18–197.

Shonkwiler, J.S., and S.T. Yen. 1999. “Two-Step Estimation of a Censored System of Equations.” *American Journal of Agricultural Economics* 81: 972–982.

Spanos, A. 1999. Probability Theory and Statistical Inference: Econometric Modeling with Observational Data. Cambridge, UK: Cambridge University Press, 1999.

Stewart, H. and S. T. Yen. 2004. “Changing Household Characteristics and the Away-from-home Food Market: a Censored Equation System Approach.” *Food Policy* 29(6):643–658.

Sznajder, M. 1999. “Ekonomia Mleczarstwa.” Wydawnictwo Akademia Rolnicza Poznań.

Sznajder, P. 2012. “Determinanty polskiego eksportu produktów mleczarskich.” *Roczniki Naukowe SERiA*, XIV (1):514–518.

Tobin J. 1958. “Estimation of Relationships for Limited Dependent Variables.” *Econometrica* 26 (1):24–36.

Toro-Vizcarrondo, C. and T.D. Wallace. 1968. “A Test of the Mean Square Error Criterion for Restrictions in Linear Regression.” *Journal of the American Statistical Association* 63(322):558–572.

Wales, T.J., and A.D. Woodland. 1983. “Estimation of Consumer Demand Systems with Binding Non-negativity Constraints.” *Journal of Econometrics* 21:263–85.

Wardle, J., Haase A.M., Steptoe A., Nillapun M., Kiriboon K., and Bellisie F. 2004. “Gender Differences in Food Choice: The Contribution of Health Beliefs and Dieting.” *Annals of Behavioral Medicine* 27(2):107–116.

Wright P. and F. Barbour. 1977. “Phased Decision Strategies: Sequels to Initial Screening.” in *Multiple Criteria Decision Making: North Holland TIMS Studies in management Science*, M. Starr and M. Zeleny, eds. Amsterdam: North-Holland Publishing Company, 91 – 109.

Wilczyński, A. 2013. “Dynamika kosztów produkcji mleka w latach 2006-2011 w wybranych krajach europejskich.” *Roczniki Naukowe SERiA*, XV (1):214–220.

Yen, S. 2005. “A Multivariate Sample-selection Model: Estimating Cigarette and Alcohol Demands with Zero Observations.” *American Journal of Agricultural Economics* 87(2) (May, 2005):453–466.

Zuehlke, T.W. and A.R. Zeman. 1991. “A Comparison of Two-Stage Estimators of Censored Regression Models.” *The Review of Economics and Statistics* 72:185–188.

APPENDIX

Table A1. Maximum likelihood estimates of Multivariate Sample-selection Model (MSSM)

Participation Equation							Level Equation						
	Whole milk			Low fat milk				Whole milk			Low fat milk		
Variable	Est.	(SE*100)		Est.	(SE*100)			Est.	(SE*100)		Est.	(SE*100)	
Constant	-0.146	(2.27)	***	0.252	(2.29)	***	Constant	1.677	(3.00)	***	1.503	(2.82)	***
Income	0.022	(0.32)	***	0.032	(0.32)	***	Income	-0.007	(0.31)	**	0.028	(0.32)	***
Educ	-0.051	(0.87)	***	-0.006	(0.88)		Educ	-0.027	(0.85)	***	0.030	(0.81)	***
Age	-0.001	(0.04)	***	0.005	(0.04)	***	Age	-0.002	(0.03)	***	0.004	(0.04)	***
Male	0.002	(0.91)		-0.086	(0.92)	***	Male	0.072	(0.88)	***	0.019	(0.87)	**
Married	0.156	(1.01)	***	0.083	(1.02)	***	Married	0.034	(1.04)	***	0.076	(1.03)	***
Employ	-0.044	(0.93)	***	0.107	(0.93)	***	Employ	-0.144	(0.89)	***	-0.029	(0.90)	***
Village	0.367	(0.89)	***	-0.460	(0.88)	***	Village	0.476	(0.96)	***	0.033	(1.10)	***
Dchild	0.182	(0.92)	***	0.062	(0.91)	***	Child	0.187	(0.42)	***	0.164	(0.46)	***
Delder	0.082	(1.15)	***	-0.055	(1.14)	***	Adult	0.164	(0.42)	***	0.113	(0.46)	***
OutD	-0.018	(0.08)	***	0.017	(0.08)	***	Elder	0.258	(0.77)	***	0.177	(0.87)	***
OutF	0.027	(0.21)	***	-0.013	(0.21)	***	OutD	-0.003	(0.07)	***	0.010	(0.08)	***
Time	0.036	(0.28)	***	-0.042	(0.28)	***	OutF	-0.015	(0.20)	***	-0.033	(0.20)	***
							Time	0.026	(0.26)	***	0.003	(0.27)	
Elements of variance-covariance matrix													
Rho.Buy1.Buy2				-0.574	(0.24)	***	Rho.Buy2.y2				0.080	(2.76)	***
Rho.Buy1.y1				0.097	(2.30)	***	Rho.y1.y2				-0.117	(1.03)	***
Rho.Buy1.y2				-0.359	(0.89)	***	Sigma.y1				0.968	(0.27)	***
Rho.Buy2.y1				-0.427	(0.73)	***	Sigma.y2				0.966	(0.25)	***
Log likelihood: -621,010.1528													

Note: *** *P*<0.01; ** *P*<0.05; * *P*<0.10

Table A2. Maximum likelihood estimates of Bivariate Two-part (B2P) Model

	Bivariate probit							Bivariate lognormal					
	Whole milk			Low fat milk				Whole milk			Low fat milk		
Variable	Est.	(SE*100)		Est.	(SE*100)			Est.	(SE*100)		Est.	(SE*100)	
Constant	-0.145	(2.277)	***	0.255	(2.290)	***	Constant	1.712	(2.360)	***	1.553	(2.293)	***
Income	0.022	(0.316)	***	0.032	(0.320)	***	Income	-0.012	(0.310)	***	0.024	(0.316)	***
Educ	-0.051	(0.866)	***	-0.004	(0.877)		Educ	-0.034	(0.862)	***	0.019	(0.826)	**
Age	-0.001	(0.037)	***	0.005	(0.037)	***	Age	-0.001	(0.034)	***	0.004	(0.035)	***
Male	0.002	(0.912)		-0.087	(0.923)	***	Male	0.067	(0.892)	***	0.021	(0.873)	***
Married	0.159	(1.006)	***	0.084	(1.023)	***	Married	0.048	(1.043)	***	0.099	(1.040)	***
Employ	-0.043	(0.932)	***	0.107	(0.934)	***	Employ	-0.145	(0.906)	***	-0.040	(0.901)	***
Village	0.366	(0.888)	***	-0.460	(0.874)	***	Village	0.486	(0.865)	***	0.091	(0.887)	***
Dchild	0.176	(0.921)	***	0.051	(0.924)	***	Child	0.197	(0.427)	***	0.172	(0.478)	***
Delder	0.085	(1.157)	***	-0.060	(1.145)	***	Adult	0.170	(0.447)	***	0.105	(0.483)	***
OutD	-0.018	(0.077)	***	0.017	(0.076)	***	Elder	0.268	(0.800)	***	0.176	(0.901)	***
OutF	0.027	(0.210)	***	-0.013	(0.213)	***	OutD	-0.001	(0.069)	*	0.011	(0.077)	***
Time	0.036	(0.276)	***	-0.042	(0.277)	***	OutF	-0.019	(0.198)	***	-0.036	(0.197)	***
							Time	0.021	(0.262)	***	0.001	(0.265)	
Rho.Buy1.Buy2				-0.578	(0.583)	***		Rho.y1.y2			-0.016	(0.556)	***
								Sigma.y1			0.964	(0.235)	***
								Sigma.y2			0.965	(0.222)	***
Log likelihood:-131,936.9541								Log likelihood: -560,965.9664					
Log likelihood with intercept only:-136741								Log likelihood with intercept only: -619,998.0294					
Log likelihood (B2P): -692,902.9205						Log likelihood (B2P) with intercept only: -756,739.0294							

Note: *** *P*<0.01; ** *P*<0.05; * *P*<0.10

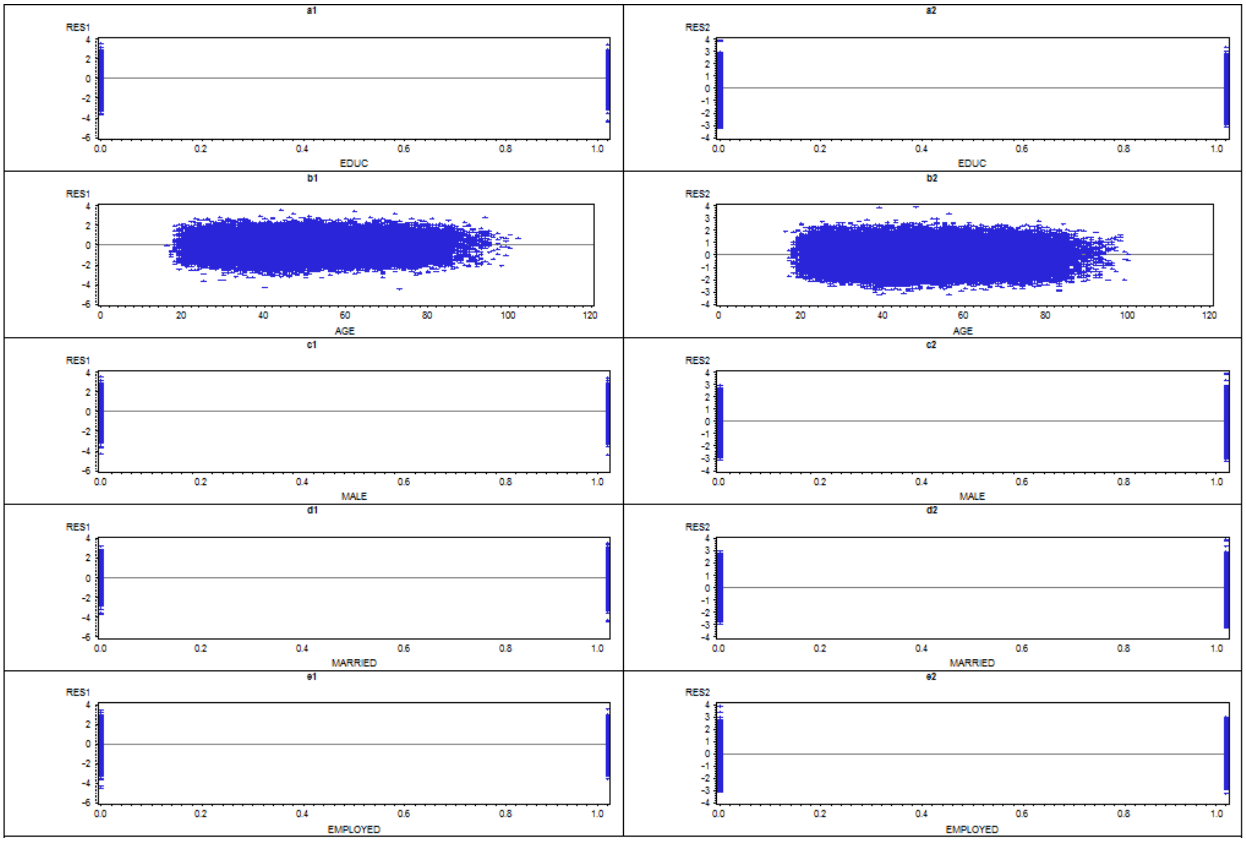

Figure A1. Residuals by Household Head Characteristics

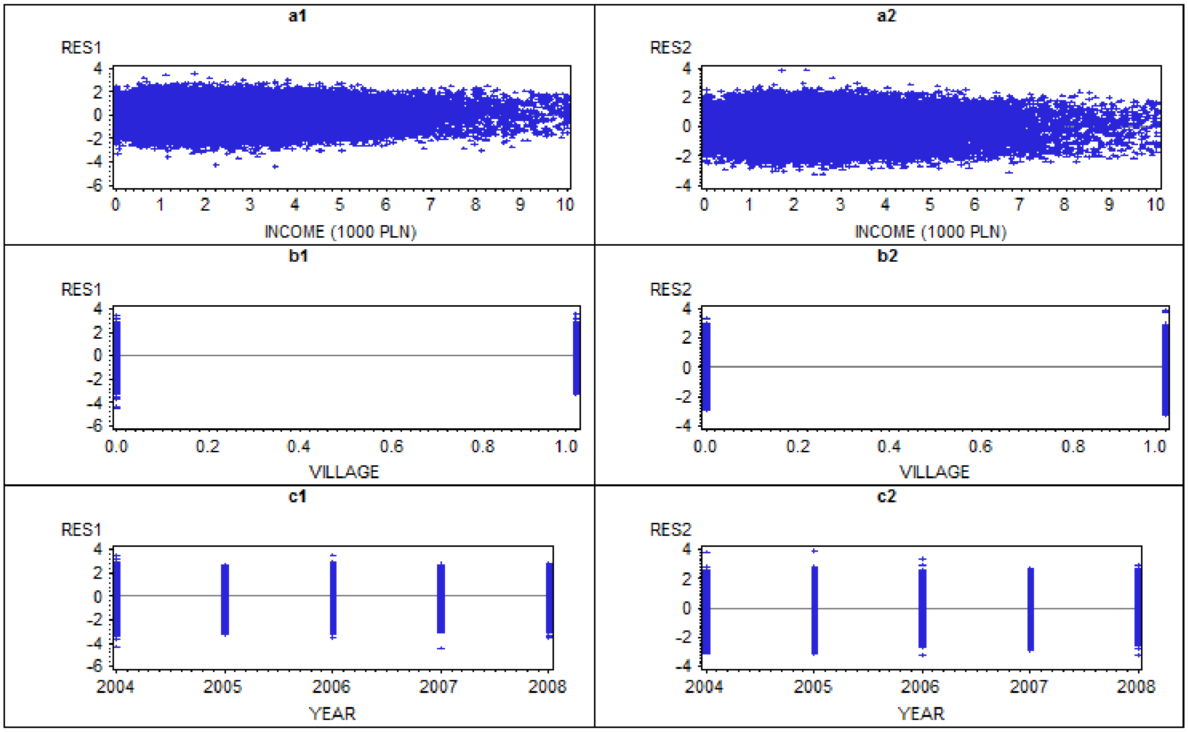

Figure A2. Residuals by Household Income, Location and Time Period

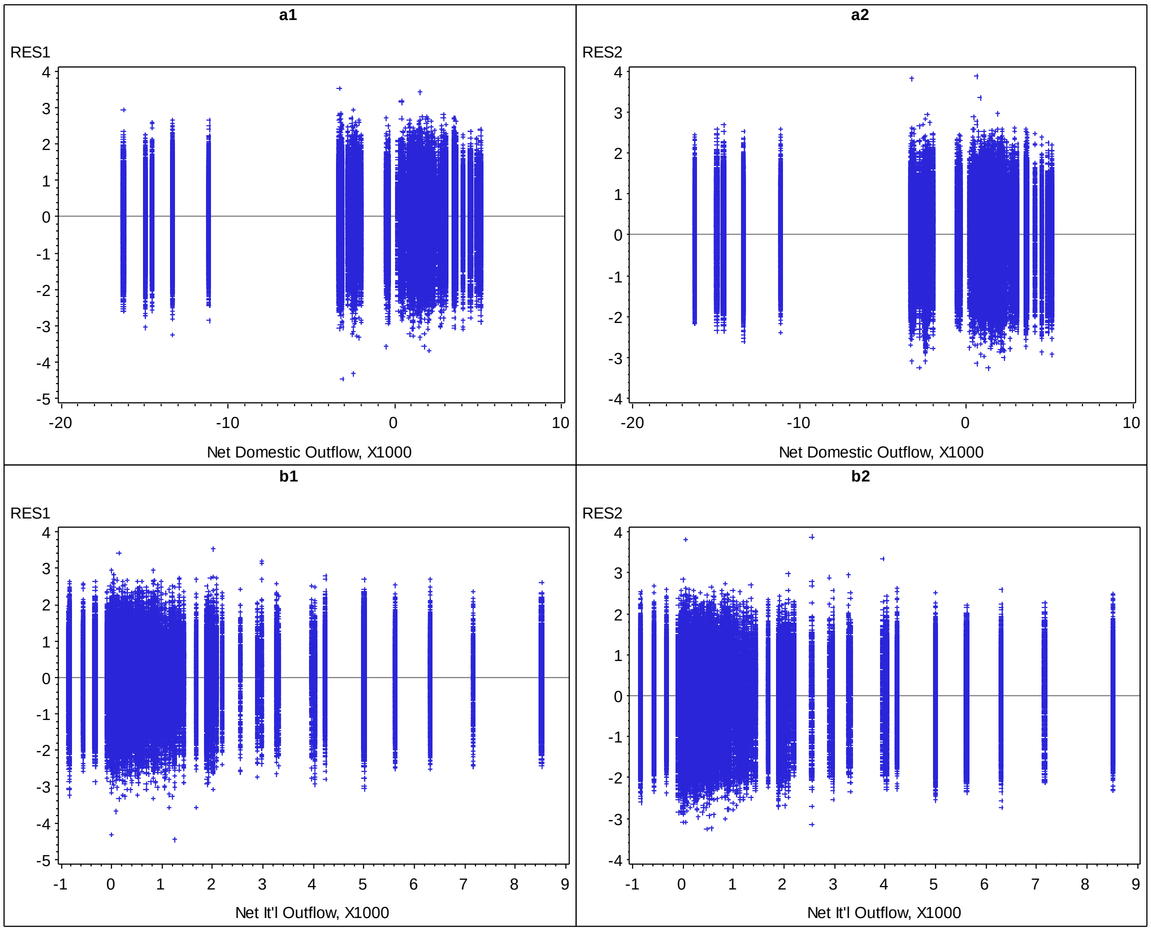

Figure A3. Residuals by Domestic and International Worker Migration

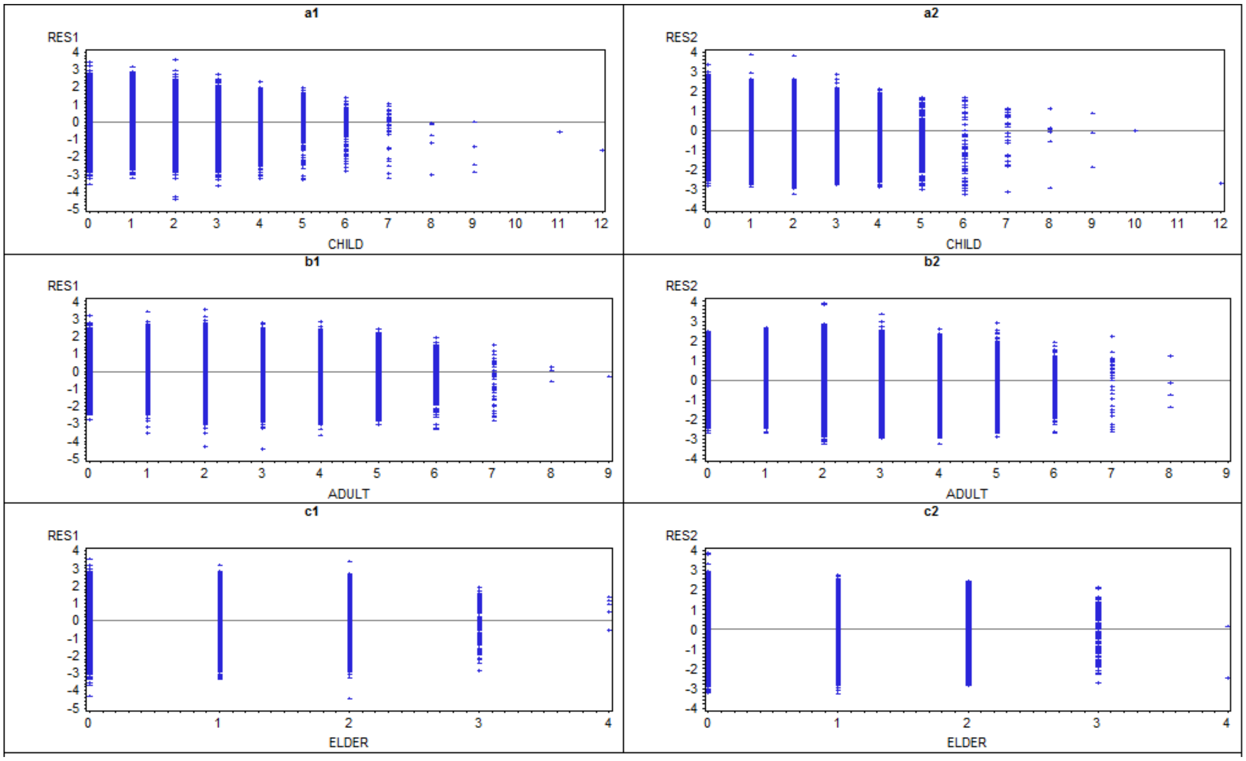

Figure A4. Residuals by Number of Family Members