Household Whole and Low Fat Milk Consumption in Poland:

A Bivariate Two-part Model

Shengfei Fu
University of Georgia
Department of Ag & Applied Economics
306 Conner Hall, Athens GA 30602
shengfei@uga.edu

Anna M. Klepacka
Warsaw University of Life Science-SGGW
Faculty of Production Engineering
164 Nowoursynowska St., 02-787 Warsaw, Poland
annaklepacka@hotmail.com

Wojciech J. Florkowski
University of Georgia
Department of Ag & Applied Economics
1109 Experiment St., 212 Stuckey Building, Griffin, GA 30223
wojciech@uga.edu

Selected Paper prepared for presentation at the Southern Agricultural Economics Association Annual Meeting, Dallas, Texas, February 1-4, 2014

Household Whole and Low Fat Milk Consumption in Poland:

A Bivariate Two-part Model

Abstract

Milk products are a preeminent food category in Poland, providing both employment and dietary benefit. This paper investigates factors affecting household milk consumption in Poland, paying attention to the effect of outmigration. Bivariate two-part model analyzes actual milk spending and allows the dependence of whole and low-fat milk consumption decisions.

Keywords: diary product, milk consumption, whole milk, low fat milk, Polish household, Poland, depopulation, outmigration, nutrition, dietary welfare, bivariate two-part model, Heckman selection model, heterogeneity, heteroskedasticity.

Introduction

Dairy products are a preeminent food category in retail sector of many countries (Sznajder, 1999). In Poland, the dairy sector is one of the most important parts of the food industry (Sznajder, 2012) and dairying is among the most important farm enterprises (Wilczynski, 2013) across Europe. The dairy sector represents about 16% of sales revenues of the Polish food processing industry (Sznajder, 2012). Milk production and consumption varies across regions in Poland with production being more unevenly distributed than consumption (Parzonko, 2013).

A single characteristic common to the leading milk consuming areas is population/worker outmigration. Leading milk consumption regions experience outmigration, depopulation and lag in economic development, due to fewer job opportunities. Because of the specificity of dairy production, expanding milk production in regions with suitable natural conditions could provide job opportunities in rural areas (Klepacka et al., 2013).

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. The investigation of demographic, socio-economic and location factors, and their connection to milk consumption are also important because of the observed substantial decline in fluid milk consumption and the milk's 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, 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. Overall, the fears of inadequate calcium intake and the underestimated role of vitamin D in health maintenance seem to outweigh potentially negative effects associated with saturated fatty acid consumption in milk. 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.

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 (Schultz1962; West and Price 1976; Benito 2006). 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. Migration leads to changes in population structure, which in return contributes to the different consumption features. Previous studies focused on the dampening effect of depopulation on the economic growth; however, less attention has been paid to the dietary welfare of people living in the depopulating regions at a micro/household level. This study investigates determinants of household expenditure on whole and low fat milk, with special attentions paid to the effect of depopulation, a current issue in Poland. The study applies quantitative methods to generate measurable effects of individual explanatory factor.

Economic Theory

A qualitative choice model based on a random utility maximization developed by McFadden (1980) provides the theoretical foundation for model specification. Our empirical model is derived by extending the discrete choice model (Pudney, 1989). 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, q is a composite commodity for other goods with price normalized to 1, 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:

$$\begin{aligned} d &= 1 \; if \; \boldsymbol{z}'\boldsymbol{\alpha} + u > 0 \qquad (2) \\ &= 0 \; if \; \boldsymbol{z}'\boldsymbol{\alpha} + u \leq 0 \end{aligned}$$

where **z** and **α** are vectors of variables and parameters affecting binary purchase decision, and u is a random error. In cross-sectional demand modeling, zero observations are often treated as the result of economic non-consumption (i.e., corner solution). 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 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:

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

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*. Denote as $\boldsymbol{x}$ the vector of income and demographic variables (with corresponding parameter vector **β**) affecting the quantity demanded.

Further, assume latent quantity y* is expressed by the lognormal distribution, which accommodates right-skewness and ensures positive purchase amount:

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

where $\boldsymbol{x}$ and $\boldsymbol{\beta}$ are variables and corresponding parameters affecting quantity decision and v is a random error.

Econometric 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), 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 existing literature to accommodate the censored data, including the Tobit model, hurdle model, two-part model, and Heckman's sample selection model. More recent development features a sample selection system or censored system in the sense of multiple-goods decisions, which allows correlation within and/or across participation decisions and intensity decisions among multiple goods. Such modeling feature is important for studying the consumption of closely related products, such as the consumption of tobacco and alcohol, and in our case, the consumption of whole and low-fat milk. A number of censored-system estimation procedures have existed 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).

However, the sample selection model and full system approach are not appropriate for the purpose of this study, because we are interested in the actual spending on milk products, rather than in the analysis of potential outcome, which involves sample selection model. Moreover, with the strong substitution effect between the consumption of whole milk and low fat milk, it is important to recognize the correlation between the consumption decisions of the two products. Therefore, a single-item two-part model is not sufficient; instead, a bivariate two-part model is used. Specifically, bivariate probit regression is used to model the decision of whether buying whole milk and/or low fat milk; bivariate lognormal regression is used to analyze positive expenditure. The bivariate two-part model can be viewed as an extension of the regular two-part model, as well as a restricted form of the full system model. Although these three models are nested in the sense of statistical form and can be tested by mean squared error (Leung and Yu 1966; Madden 2008), the theoretical and primary reason for choosing two-part model is the interest in the actual outcome.

Previous applications of two-part model usually made the assumption of homoscedasticity, and thus failed to accommodate heteroskedasticity, a common feature of cross sectional data due to heterogeneity of cross sectional unit. We specify a bivariate two-part model, accommodating heteroskedasticity.

Each outcome variable y_i (milk expenditure) is governed by a binary selection rule of whether to consume:

$$log(y_i) = \boldsymbol{x}'\boldsymbol{\beta_i} + v_i \; if \; \boldsymbol{z}'\boldsymbol{\alpha_i} + u_i > 0$$

$$y_i = 0 \; if z'\alpha_i + u_i \leq 0 \quad i = 1,2 \qquad (5)$$

Specifically, the participation decision, buying milk or not, is modeled as a bivariate probit regression:

$$\mathrm{d_i} = 1 \text{ if } \boldsymbol{z}'\boldsymbol{\alpha_i} + u_i > 0$$

$$= 0 \text{ if } \boldsymbol{z}'\boldsymbol{\alpha_i} + u_i \leq 0, \qquad \mathrm{i} = 1,2 \quad (6)$$

where disturbance terms u_1 and u_2 are randomly distributed with bivariate normal regression with zero means and covariance matirx $\mathbf{\Sigma_1} = \begin{bmatrix} \sigma_{u_1}^2 & \sigma_{u_1}\sigma_{u_2}\rho_u \\ \sigma_{u_1}\sigma_{u_2}\rho_u & \sigma_{u_2}^2 \end{bmatrix}$ and the standard deviations $\sigma_{u_1}$and $\sigma_{u_2}$are allowed to vary across observations.

The consumption decision is modeled as a bivariate lognormal regression:

$$log(y_i) = \boldsymbol{x'\beta_i} + v_i \, , i = 1,2 \quad (7)$$

where the disturbance terms v_1 and v_2 are distributed with a bivariate normal distribution of zero means and covariance matrix $\mathbf{\Sigma_2} = \begin{bmatrix} \sigma_{v_1}^2 & \sigma_{v_1}\sigma_{v_2}\rho_v \\ \sigma_{v_1}\sigma_{v_2}\rho_u & \sigma_{v_2}^2 \end{bmatrix}$ and the standard deviations σ_{v_1} and σ_{v_2} also vary across observations.

For the nth observation, to facilitate presentation of the log likelihood function, define error term standard deviations as $\mathbf{S_u} = \mathrm{diag}(\sigma_{u_1}, \sigma_{u_2})$ and $\mathbf{S_v} = \mathrm{diag}(\sigma_{v_1}, \sigma_{v_2})$. The subscript n is omitted for notational convenience, but note the error term standard deviations vary across observations. Let $\mathbf{R_{uu}} = [\rho_{\boldsymbol{ij}}^{\mathrm{uu}}]$, and $\mathbf{R_{vv}} = [\rho_{\boldsymbol{ij}}^{\mathrm{vv}}]$ be 2 x 2 correlation matrices among elements of $\boldsymbol{u}$ and $\boldsymbol{u}$, and $\boldsymbol{v}$ and $\boldsymbol{v}$, respectively. Therefore $\mathbf{\Sigma_1} = \mathrm{E}(\boldsymbol{u'u}) = \mathbf{S_u}' * \mathbf{R_{uu}} * \mathbf{S_u}$, and $\mathbf{\Sigma_2} = \mathrm{E}(\boldsymbol{v'v}) = \mathbf{S_v}' * \mathbf{R_{vv}} * \mathbf{S_v}$. Note the standard deviations are allowed to vary across observations and thus heteroskedasticity is modeled in both participation and intensity equations.

To construct the log likelihood function for the bivariate probit regression, the whole sample is decomposed into three regimes. The likelihood for positive regime (where $\mathrm{d}_1 = 1$, $\mathrm{d}_2 = 1$) is:

$$\mathrm{L1}_1 = \int_{\boldsymbol{u}>-\mathbf{r}}^{\infty} \phi(\boldsymbol{u})\mathrm{d}\boldsymbol{u} \qquad (8).$$

where the probability density function (pdf) $\phi(\boldsymbol{u}) = (2\pi)^{-1}|\Sigma_1|^{-1/2}\mathrm{e}^{-\frac{1}{2}\boldsymbol{u}'\boldsymbol{\Sigma_1}^{-1}\boldsymbol{u}}$ and $\mathbf{r} = \begin{pmatrix}\boldsymbol{z}'\boldsymbol{\alpha_1}\\ \boldsymbol{z}'\boldsymbol{\alpha_2}\end{pmatrix}$.

The likelihood for negative regime (where $\mathrm{d}_1 = 0$, $\mathrm{d}_2 = 0$) is:

$$\mathrm{L1}_2 = \int_{-\infty}^{\boldsymbol{u}\leq -\mathrm{r}} \phi(\boldsymbol{u})\mathrm{d}\boldsymbol{u} \qquad (9).$$

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:

$$\mathrm{L1}_3 = \int_{\boldsymbol{u_i}>\boldsymbol{r_i}}^{\infty}\int_{-\infty}^{\boldsymbol{u_j}\leq \boldsymbol{r_j}} \phi(\boldsymbol{u})\mathrm{d}\boldsymbol{u_j}\mathrm{d}\boldsymbol{u_i} \qquad (10).$$

The sample likelihood function is the product of $\mathrm{L1}_1$, $\mathrm{L1}_2$ or $\mathrm{L1}_3$ across observations, depending on the regimes of each observation.

For the bivariate lognormal regression, the likelihood function is: $\mathrm{L2} = (2\pi)^{-1}|\Sigma_2|^{-1/2}\mathrm{e}^{-\frac{1}{2}\boldsymbol{v}'\boldsymbol{\Sigma_2}^{-1}\boldsymbol{v}}$. Products across observations produce the likelihood function for the whole sample of positive expenditure.

Data, sample and variables

The data are from the Polish household panel of about 20,000 households annually surveyed by Poland's National Statistics 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 108,747 observations with non-missing values for the period of 2004 to 2008.

The dependent variables are expenditures in the month preceding survey on whole and low fat milk. Positive expenditures are logarithm transformed to mitigate deviation from normality and heteroskedasticity.

Two variables are reported as measure of depopulation. First, net domestic migration measures the net outflow of population from a region to other regions within Poland. Second, net international migration measures the net outflow of population from a region to other countries.

Other demographic and socio-economic factors 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).

Table 1 presents summary of statistics of sample variables. Rural residents accounts for 36.1% of all observed households (Village=1). Household income in the month preceding survey averages at 2383 Polish Zloty (PLN). Slightly more than 60% households are headed by male members. And, 67.9% of household heads are married. The proportion of household heads with secondary or higher education is 44.9%. The average household head's age is 50.8 years. In term 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, 1.83 and 0.45 per household, respectively. Net outmigration population from a region averages at 1328.6 persons over all 16 administrative regions of Poland. About 30% households are observed in 2004 and about 17%~18% households are observed in each year of 2005 to 2008. Lastly, the percent of households who bought whole and low fat milk in the month preceding survey is 61.2% and 62.4%, respectively. Conditional on purchase, households on average spend PLN19.64 per month on whole milk and PLN14.45 on low fat milk.

Results

As shown in Table 2, parameter estimates are obtained by maximum likelihood estimation. The left part of Table 2 reports the maximum likelihood estimates of bivariate probit regression for buying whole and low fat milk and the right part shows the parameter estimates of bivariate lognormal regression for expenditure on the two milk products.

The correlation parameters in both equations are estimated to be negative and statistically significant at 5% level. This confirms our expectation and validates the use of bivariate regression in both decision equations. As expected, the decision to buy whole milk is negatively correlated with the decision to buy low fat milk. Also, the expenditure on whole milk is negatively correlated with the expenditure on low fat milk.

Exponential link of error standard deviation is used to accommodate heteroskedasticity. Households with different size and income usually have very different consumption patterns (Nelson 1988; Deaton and Paxson 1998). Therefore, we use the variables income and number of family members to model heteroskedasticity. Households with higher income level have fewer constraints on food choice and, therefore, are expected to have higher variance in the decision of buying milk. As family size increases, it is more likely that at least one member wants to consume milk. Consistent with this expectation, bigger-sized households are estimated to have lower variances in the decision to buy milk. Higher variances of both milk expenditures are estimated for households with higher income and larger size.

The signs of all variables expect for marital status differ across the participation and intensity equation. The result implies different decision rules are applied when households decide whether to buy milk and if they buy, how much to spend. Thus, such behavior mandates the use of two different equations.

For most demographic variables, the signs of parameter estimates are consistent with expectations and previous results reported in literature. Married household heads have higher probabilities of buying both milk products and also spend more on both products.

In the binary decision to buy milk, income positively influences the probability of buying whole and low fat milk. *Ceteris paribus*, more affluent households spend more on low fat milk, but less on whole milk.

Household heads with higher education have lower probability buying whole milk and are more likely to buy low fat milk, compared to their counterparts with lower educational attainment level. Low fat milk has the same nutritional benefits as whole milk but contains less fat and is generally considered healthier. People with higher education are more likely to recognize the relative superiority of low fat milk and, therefore, are more likely to buy it. However, higher education has a negative effect on the expenditure of both milk products, possibly because of wider food choices by people with higher education.

The effect of more family members (children, adults, or elders, respectively) is estimated to increase the probability of buying whole milk, but decrease the probability of buying low fat milk. Although the effects of these three measures are of the same sign, their values vary, especially for low fat milk consumption. 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. More children and elders, respectively, increase the average expenditure on both whole and low fat milk. However, conditional on purchase, more adults are associated with higher expenditure on low fat milk and lower expenditure on whole milk.

Older household heads are associated with higher expenditure on both milk products, as older population generally consumes more dairy products. In addition, elder household heads are more likely to buy low fat milk and less likely to buy whole milk. this is because older people are more concerned about dietary healthiness the proportion of people with lactose intolerance increases among older population.

The variable male household head is insignificant regarding consumption decision of whole milk. But male household head are less likely to buy low fat milk. 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 low fat milk, male household heads on average make larger purchase.

Households with relatively stable employment are less likely to buy whole milk, while more likely to buy low fat milk. And employment stability is positively related to higher milk expenditures.

Households residing in villages have higher probability of buying whole milk and lower probability of buying low fat milk, reflecting relatively inferior milk choice in rural area. Conditional on purchase, however, rural residents spend more on both milk products, compared to their urban counterparts.

Last but not the least, depopulation/outmigration shows interesting effects. Higher outmigration is associated with lower probability of buying whole milk, higher probability of buying low fat milk, and higher expenditure on both products. The possible underlying mechanism 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. Unfortunately, our data set does not facilitate the empirical investigation of such mechanism. Since regions with high outmigration tend to be economically less developed, the supported development of diary sector (at least milk products) in these areas may provide the potential of additional employment and economic development.

Discussions and Conclusions

Dairy products, mainly milk products, are a preeminent food category in the retail sector of Poland. The dairy sector represents about 16% of sales revenue of the Polish processing industry (Sznajder, 2012). Milk is 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 investigated 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). An additional variable, depopulation, measures the net outmigration from a region.

Once considering the actual spending on whole and low fat milk, two-part model was chosen over sample selection model. In order to model the substitution effect between the consumption of whole and low fat milk, we used a bivariate two-part model and modeled heteroskedasticity associated with family income and family size.

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 and households with married heads have higher probabilities buying both milk products. Higher education, older household heads, stable employment, and higher depopulation level are associated with relatively healthier milk choice as they increase the probability of buying low fat milk, while lowering the likelihood of whole milk purchase.

However, larger-sized households and rural residents usually appear to make less healthy fluid milk choice. Households with more children, adults, and elders, and rural residents are more likely to buy whole milk and less likely to buy low fat milk. Male household heads are also less likely to buy low fat milk.

Conditional on purchase, above variables are found to have different effects on the intensity decision than they do on the participation decision. More children and elders, older, male, or married household heads, more stable employment, and higher level of depopulation are associated with higher expenditures on both milk products.

People with higher education level and residing in villages spend less on both products. Households with higher education and more adults spend more on low fat milk and buy less whole milk.

In summary, the bivariate two-part model enables a close scrutiny on 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. The estimated effect of depopulation implies the importance of fluid milk in diets of regions loosing residents. Also, the study provides a notion about potential of diary sector (milk products) for local employment and economic development.

References

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

Bettman, J.R. 1979. "Memory Factors in Consumer Choice: A Review." *Journal of Marketing* 43(Spring 1979): 37-53.

Deaton, A. and C. Paxson. 1998. "Economies of scale, household size, and the demand for food."*Journal of political economy* 106(5) (October 1998):897-930.

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

Greene, W.H. 2012. *Econometric Analysis*, 7th. ed. New York: Pearson Addison Wesley.

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

GUS (Central Statistical Office) 2010. Budzety gospodarstw domowych 2009. Warszawa.

Halvorsen R. and R. Palmquist. 1980. "The Interpretation of Dummy Variables in Semilogarithmic Equations." *American Economic Review* 70(3) (Jun., 1980):474-475.

Heckman J.J. 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 & Economic Statistics* 8(3): 365-371.

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.

Koloszyc, E. 2013. Dochodowość produkcji mleka w gospodarstwach na świecie w 2012 roku. *Roczniki Naukowe SERiA,* XV(2):165-170.

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

Leung, S.F. and S. Yu. 1996. "On the choice between sample selection and two-part models." *Journal of Econometrics* 72: 197–229.

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.

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

Nelson, J. 1988. "Household economies of scale in consumption: theory and evidence." *Econometrica* 56(6) (November 1988):1301-1314.

Orłowska, A. 2012. Zachowania nabywcze konsumentów produktów mleczarskich. *Roczniki Naukowe SERiA*, XIV(1):353-357.

Parzonko, A. 2013. Regionalne zróżnicowanie produkcji mleka w Polsce – uwarunkowania przyrodnicze i ekonomiczne. *Roczniki Naukowe SERiA*, XV(2):265-270.

Perali, F., and J.P. Chavas. 2000. "Estimation of Censored Demand Equations from Large Cross-Section Data." *American Journal of Agricultural Economics*82:1022–37.

Pudney, S. 1989. "Modelling Individual Choice: The Econometrics of Corners, Kinks, and Holes." Cambridge, UK: Blackwell Publishers.

Shonkwiler, J.S., and S.T. Yen. 1999. "Two-Step Estimation of a Censored System of Equations." *American Journal of Agricultural Economics* 81:972–82.

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): 181-197.

Sobczynski, T. S., A. M. Klepacka, C. Revoredo-Giha, and W. J. Florkowski. 2013. The Differences in Cost Efficiency of Dairy Farms in Four Regions of Poland. Paper presented at the International Farm Management Association, Warszawa, July 21-26, 2013, II (52):114-122.

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

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

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.

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

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

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.

Wysokiński, M. 2012. Produkcja mleka w Polsce na tle Europy i Świata. *Roczniki Naukowe SERiA*, XIV (2):201-206.

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.

Table 1. Summary of Descriptive Statistics of Sample Variables

Variable	Description/Unit	Mean	Std Dev
Demographic, Socio-Economic Factors / Explanatory variables			
Village	1, if a household residents in village, 0 otherwise	0.361	0.480
Income	Household income in the month preceding survey, in 1000 Polish Zloty (PLN)	2.383	1.886
Male	1, if the household head is male, 0 otherwise	0.596	0.491
Married	1, if the household head is married, 0 otherwise	0.679	0.467
HighEduc	1, if the household head has secondary or higher education, 0 otherwise	0.449	0.695
Age	Household head's age, in years	50.847	15.272
Employed	1 if household head is permanently employed or contract employee, 0 otherwise	0.421	0.494
Children	Number of children (under 18)	0.639	0.984
Adult	Number of adults 60 or under 60 years old	1.827	1.198
Elder	Number of elders above 60	0.449	0.695
OUTD	Net migration domestically to other regions in Poland, in 1000	-1.283	5.621
OUTF	Net migration international to other countries, in 1000	1.3286	2.051
Depop	Net migration, OUTD+OUTF		
Year2004	Baseline, 1 if observed in 2004, 0 otherwise		
Year2005	1 if observed in 2005, 0 otherwise	0.178	0.383
Year2006	1 if observed in 2006, 0 otherwise	0.178	0.382
Year2007	1 if observed in 2007, 0 otherwise	0.176	0.381
Year2008	1 if observed in 2008, 0 otherwise	0.176	0.381
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 PLN	19.644	22.208
Lmilk	Expenditure on low fat milk in the month preceding survey, in PLN	14.453	14.716

Table 2. Maximum-likelihood estimates of bivariate two-part model for milk consumption

Bivariate probit for buying whole and low fat milk

	Whole milk			Low fat milk		
Variable	Coeff. Est. (Std. Err.)			Coeff. Est. (Std. Err.)		
Intercept	-0.379	(0.037)	***	0.232	(0.022)	***
Income	0.131	(0.012)	***	0.132	(0.008)	***
HighEduc	-0.059	(0.013)	***	0.004	(0.009)	
Children	0.117	(0.007)	***	-0.028	(0.003)	***
Adult	0.114	(0.007)	***	-0.059	(0.004)	***
Elder	0.180	(0.014)	***	-0.113	(0.009)	***
Age	-0.002	(0.001)	***	0.004	(0.000)	***
Male	-0.012	(0.013)		-0.075	(0.009)	***
Married	0.104	(0.015)	***	0.083	(0.011)	***
Employed	-0.103	(0.015)	***	0.081	(0.009)	***
Village	0.477	(0.017)	***	-0.415	(0.013)	***
Depop	-0.015	(0.001)	***	0.010	(0.001)	***
Year2005	0.053	(0.016)	***	-0.067	(0.011)	***
Year2006	0.207	(0.017)	***	0.001	(-0.101)	***
Year2007	0.218	(0.018)	***	-0.139	(0.012)	***
Year2008	0.226	(0.018)	***	-0.195	(0.013)	***
Rho	-0.579	(0.004)	***			
H.Intercept	--					
H.Income	0.361	(0.024)	***	0.434	(0.017)	***
H.NumFm	--			-0.297	(0.011)	***
Number of Observation 108747				Log Likelihood -132328		

Bivariate lognormal regression for expenditure on whole and low fat milk

	Whole milk			Low fat milk		
Variable	Coeff. Est. (Std. Err.)			Coeff. Est. (Std. Err.)		
Intercept	1.673	(0.023)	***	1.478	(0.022)	***
Income	-0.007	(0.003)	**	0.013	(0.002)	***
HighEduc	-0.011	(0.002)	***	-0.022	(0.003)	***
Children	0.237	(0.009)	***	0.247	(0.009)	***
Adult	-0.034	(0.009)	***	0.025	(0.008)	***
Elder	0.172	(0.004)	***	0.117	(0.005)	***
Age	0.272	(0.008)	***	0.196	(0.009)	***
Male	0.000	(0.000)		0.004	(0.000)	***
Married	0.065	(0.009)	***	0.020	(0.009)	**
Employed	0.039	(0.010)	***	0.092	(0.010)	***
Village	-0.139	(0.009)	***	-0.047	(0.009)	***
Depop	0.501	(0.009)	***	0.111	(0.009)	***
Year2005	0.010	(0.012)		-0.020	(0.011)	*
Year2006	-0.025	(0.011)	**	-0.073	(0.011)	***
Year2007	0.026	(0.011)	**	-0.037	(0.011)	***
Year2008	0.082	(0.011)	***	0.006	(0.012)	
Rho	-0.012	(0.006)	**			
H.Intercept	-0.152	(0.006)	***	-0.168	(0.005)	***
H.Income	0.030	(0.002)	***	0.042	(0.002)	***
H.NumFm	0.007	(0.001)	***	0.003	(0.001)	**
Number of Observation 102305				Log Likelihood -250157		

* significant at 10% ** significant at 5% *** significant at 1%
Exponential link for error term standard deviation