

Smutna, Sarlota; Scasny, Milan

Working Paper

Selectivity Problem in Demand Analysis: Single Equation Approach

IES Working Paper, No. 21/2017

Provided in Cooperation with:

Charles University, Institute of Economic Studies (IES)

Suggested Citation: Smutna, Sarlota; Scasny, Milan (2017) : Selectivity Problem in Demand Analysis: Single Equation Approach, IES Working Paper, No. 21/2017, Charles University in Prague, Institute of Economic Studies (IES), Prague

This Version is available at:

<https://hdl.handle.net/10419/174214>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Institute of Economic Studies, Faculty of Social Sciences
Charles University in Prague

Selectivity Problem in Demand Analysis: Single Equation Approach

Sarlota Smutna
Milan Scasny

IES Working Paper: 21/2017



Institute of Economic Studies,
Faculty of Social Sciences,
Charles University in Prague

[UK FSV – IES]

Opletalova 26
CZ-110 00, Prague
E-mail : ies@fsv.cuni.cz
<http://ies.fsv.cuni.cz>

Institut ekonomických studií
Fakulta sociálních věd
Univerzita Karlova v Praze

Opletalova 26
110 00 Praha 1

E-mail : ies@fsv.cuni.cz
<http://ies.fsv.cuni.cz>

Disclaimer: The IES Working Papers is an online paper series for works by the faculty and students of the Institute of Economic Studies, Faculty of Social Sciences, Charles University in Prague, Czech Republic. The papers are peer reviewed, but they are *not* edited or formatted by the editors. The views expressed in documents served by this site do not reflect the views of the IES or any other Charles University Department. They are the sole property of the respective authors. Additional info at: ies@fsv.cuni.cz

Copyright Notice: Although all documents published by the IES are provided without charge, they are licensed for personal, academic or educational use. All rights are reserved by the authors.

Citations: All references to documents served by this site must be appropriately cited.

Bibliographic information:

Smutna S., Scasny M. (2017). " Selectivity Problem in Demand Analysis: Single Equation Approach"
IES Working Paper 21/2017. IES FSV. Charles University.

This paper can be downloaded at: <http://ies.fsv.cuni.cz>

Selectivity Problem in Demand Analysis: Single Equation Approach

Sarlota Smutna^{a,b}

Milan Scasny^{a,b}

^aInstitute of Economic Studies, Faculty of Social Sciences, Charles University
Opletalova 21, 110 00, Prague, Czech Republic

^bCharles University Environment Centre, José Martího 407/2, 162 00, Prague,
Czech Republic

Email (corresponding author): sarlota.smutna@gmail.com

September 2017

Abstract:

This paper deals with a problem of censored data in the household demand analysis when budget survey data is used. Micro-data, in contrast with aggregated data, usually contains a significant portion of zero observations (no consumption recorded) that leads to censoring of data and potential selectivity problem resulting in biased estimates if inappropriate econometric model is used. We review different treatment methods available in the literature that control the selectivity problem. Concretely, it is Tobit model, Two-part model, Double-hurdle model, Sample selection model with three different estimators – FIML, Heckman two-step, and Cosslett's semi-parametric estimator. On the empirical example we indeed show that firstly the treatment methods are necessary also for small levels of censoring and secondly the choice of treatment method matters even for different products within the same dataset. We compare performance over the above single-equation demand models together with OLS. The household demand is analysed for 13 different food products with high variety of level of censoring. We found that the Heckman two-step procedure and Cosslett's semi-parametric estimators performed best among all examined techniques in our case and that these two estimators yield similar estimates of income and own price elasticities. The Two-part model performs equivalently but the estimation results differ from the Heckman two-step and the Cosslett's estimator. The OLS estimates are biased and perform poorly together with Tobit model with weak performance.

Keywords: demand analysis, censoring, selectivity, Heckman two-step estimator

JEL: C24, D12, R22

1 Introduction

Consumer demand analysis is an important topic in applied econometrics. In general, elasticity estimates are the most applicable outcome of demand analyses that help to explain the sensitivity of consumer towards the changes in prices of goods and/or consumer income. It is therefore of high importance to obtain unbiased estimates of the key demand parameters. To an increasing extent, the demand analyses rely on individual-level or household-level expenditure survey data describing consumption patterns of each surveyed individual, or household, respectively. This is also the only type of data containing the socio-demographic, attitudinal, and other information about consumers that allows to analyze observed preference heterogeneity in consumer demand.¹ On the other hand, this data may contain considerably large share of zero observations (i.e. no consumption recorded) due to the different reasons such as infrequency of purchase, for instance.

Such incompleteness of observations often leads to non random, selected samples that rise the selectivity problem. In general, the data incompleteness is usually formed by censoring or truncation of the data. Truncation means that the whole information, i.e. the dependent variable and the regressors, is missing. It can be simply caused by the fact that some specific group of population was excluded or not reached in the survey (for example, on-line buyers if a survey is carried out in a shop). Censoring of data means that the information on the dependent variable is only lost, for example, if there is a threshold. In this paper, we pay attention to the censored data. Then in case of selectivity problem, common estimators such as Ordinary Least Squares (OLS) become biased. The selectivity problem has to be treated in order to obtain unbiased results of the analysis. Different treatment methods have been developed over time. The Heckman two-step estimator is well known, but not only one, treatment method in the case of the single equation demand analysis.

This paper aims to review different treatment methods for single equation framework and to illustrate their use and performance on the empirical example. Despite the frequent use of consumer expenditure survey data and hence frequent use of zero treatment methods, systematic comparison of these methods is missing in the literature. Moreover, some of the treatment methods are used in the analyses without any discussion about its suitability. This paper aims to fill this gap and to point out the importance of the appropriateness of employed treatment method.

Specifically, we present six different techniques to treat the selectivity in the single-equation demand model. These models are tobit, double-hurdle, two part and sample selection model with three different estimators - FIML, LIML (Heckman two step estimator), and semi-parametric Cosslett's estimator. To show that the treatment methods are necessary even for a small portion of zeros in the sample and that the choice about the treatment method matters, we use these approaches, alongside with OLS, to estimate household demand for 13 specific food items. These food items differ with respect to percentage of zeros (no consumption) between 3 % (milk) and 72 % (cocoa). We examine statistical performance of each of the six treatment techniques (compared to untreated OLS) in order to decide which one is performing the best in each case.

For our empirical application we choose the demand for different food items as they provide us a range of products with different levels of censoring. Beside the distribution and correlation of disturbances in the participation (to buy or not to buy) and outcome equation (how much to buy) and other factors, the level of censoring is one of the characteristics considered in performance

¹Scanner data, collected from the electronic counters in stores may contain also a limited portion of sociodemographic information, if stores offer registrations to their customers in order to gain more benefits or discounts.

studies discussed later (the Monte Carlo simulation studies). Furthermore, the level of censoring is easily observable. That is why the varying level of censoring was the main decisive factor for our choice of products' category. In fact, different food products present a great variety of levels of censoring (from 0.27 % to 84.70 % in our dataset) comparing to energies or transport, for instance, where no such a big variety can be observed.² Thanks to the different types of food commodities we are also able to examine whether even within one dataset and one good (food) category the best performing treatment method differ

The paper is further structured as follows. Section 2 gives an overview on censoring data problems and treatment methods. Section 3 describes the methodology. Following sections bring the data description and model specification and Section 6 presents the results together with their discussion. Last section concludes.

2 Treatment models: A Review

In the demand analysis when using budget survey data we observe either real positive numbers that stand for quantity purchased/consumed or for expenditures for analysed item or zeros. Such structure type of data is called left-censored as the dependent variable is bounded from the left but the other variables are not influenced. Denoting by y^* completely observed (latent) dependent variable and by y actually observed variable (the one that is at researcher's disposition), the censoring can be illustrated as:

$$y = \begin{cases} y^* & y^* > T \\ - & y^* \leq T \end{cases} \quad (1)$$

with T as a threshold at the left. In the demand analysis with zero observations, we rewrite the problem as

$$y = \begin{cases} y^* & y^* > 0 \\ 0 & y^* \leq 0. \end{cases} \quad (2)$$

Basic model that deals with censoring of data is the tobit model proposed by Tobin (1958). After probit model that is used to explain zeros and ones as dependent variables Tobin (1958) has built a model with dependent variable that has only strictly positive values. Moreover, his model is the linear regression model, thus

$$y^* = X'\beta + \epsilon \quad (3)$$

$$y = \begin{cases} y^* & y^* > 0 \\ - & y^* \leq 0 \end{cases} \quad (4)$$

²While the proportion of zeros considerably vary across the food product categories, the share of households, for example, that do not consume certain energy item is conditional on energy infrastructure, connection to electricity, gas, or district heating utility. The consumption of certain goods can be also dependent on availability (for instance, public transport). As a consequence, the share of zeros does not vary so much as in the case of food items. For instance, there are 35.8 % of Czech households not using gas, 53.8 % not using trains, or 43.7 % not using intercity public transport by buses.

where ϵ are disturbances that can be normally distributed. The theory behind supposes that the same stochastic mechanism determines the probability of limit response (the probability that $y^* > 0$) and its size. For demand analysis, it practically means that tobit model assumes the same stochastic process (regressors) influences the decision of purchase and the quantity purchased. The author proposed firstly this model for demand for durable goods such as automobiles. Assuming the distribution of disturbances, usually normal distribution, the model can be estimated by MLE. An alternative estimator is for example trimmed least squares (Powell, 1986). Amemiya (1974) developed tobit model further for a system of equations that is commonly called Tobin-Amemiya model.

In the demand analysis, it is very unlike that the same parameters influence in the same magnitude the purchase and quantity decision. For example, the decision about purchase/ consumption can be induced by personal convictions (transport pollution, vegetarianism), by religious or cultural convictions (food and Judaism), dietary restrictions, unavailability or no possibility to purchase/consume (public transport does not cover all villages). These factors will not then influence the quantity purchased by the other consumers.

Hence, two specific mechanisms are needed, one that influences the choice of product to purchase and the second one that describes the decision of quantity afterwards. Distinguishing these two processes we deal with the decision made in two steps modelled by two equations. Together with censoring, the problem can be illustrated for the demand analysis as follows:

$$y_1^* = X_1' \beta_1 + u \quad (5)$$

$$y_2^* = X_2' \beta_2 + v \quad (6)$$

$$y_2 = \begin{cases} y_2^* & y_1^* > 0 \\ 0 & y_1^* \leq 0. \end{cases} \quad (7)$$

The equation (5) is called participation equation (participation on the market of given product), the regressors X_1 influence the decision of purchase and y_1^* is usually a dummy variable indicating purchase or non purchase. The following equation (6) is called outcome equation and X_2 are regressors that determine the quantity of purchased item. The u and v are disturbances that can be correlated.

In case the disturbances of two equations are independent, the double-hurdle or two-part model are defined. The double-hurdle model was proposed by Cragg (1971). In the notation of Jones (1989), the model can be rewritten as:

$$\text{participation equation: } y_1^* = X_1' \beta_1 + u \quad (8)$$

$$\text{outcome equation: } y_2^* = X_2' \beta_2 + v \quad (9)$$

$$\text{observation: } y_2 = \begin{cases} X_2' \beta_2 + v & y_1^* > 0 \wedge y_2^* > 0 \\ 0 & \text{otherwise.} \end{cases} \quad (10)$$

The name of the model arises from the fact that in order to observe positive quantity demanded two hurdles has to be overcome. Firstly, the consumer has to be interested in the market of the product and secondly she has to buy some quantity of the product in the given moment. It means that this model allows for the infrequency of purchase. This stands for the situation when the consumer is interested in the product and she usually buys it but in the moment of observation

there was no need to buy it as it is long lasting item or its purchase is not necessary so often. We can imagine durable goods such as automobiles, or consumer that usually travels by plane but there was no journey needed in the given period of time. Under the assumption of distribution of error terms, the whole model, coefficients of both equations can be estimated by MLE.

Another model that assumes no correlation of disturbances u and v is two-part model. Cragg (1971) is considered to be the first who proposed this model as well. We consider the participation equation in the usual way but for the outcome equation, only the positive observation are allowed. It means that the zero observations are not included in the estimation. The model looks as follows:

$$y_1^* = X_1' \beta_1 + u \quad (11)$$

$$y_2 | y_1^* > 0 = X_2' \beta_2 + v. \quad (12)$$

It is possible to estimate the whole model by MLE such as in Jones and Yen (2000). As the second equation is defined conditionally, it is also possible to estimate first equation as a probit model and use for the second equation OLS estimator on positive observations only.

In case that the error terms u and v are correlated, the model defined by equations (5), (6), and (7) is called sample selection model and there exist different methods to estimate it. This model nests the model with independent error terms and it is very often used in consumer demand (Puhani, 2000). Assuming the bivariate normal distribution of error terms, the (log) likelihood function can be determined and the model can be estimated by MLE such as in Leung and Yu (1996). This form of estimation is sometimes called in the literature as full-information maximum likelihood (FIML).

Another estimator of sample selection model is Heckman two-step estimator. It can be also used under the name limited-information maximum likelihood (LIML) estimator. It was proposed and developed by Heckman (1976, 1979). At that time, the MLE was computationally demanding that is why Heckman derived a simpler method to estimate the sample selection model. Despite the fact that FIML is nowadays easy to estimate, the two-step LIML estimator is still preferred in empirical studies. Heckman derived the bias of usual OLS when it is applied as an estimator in the sample selection problem. The bias is defined as:

$$E[y_2 | X_2, y_1^* > 0] = X_2 \beta_2 + E[v | X_2, y_1^* > 0], \quad (13)$$

where

$$\begin{aligned} E[v | X_2, y_1^* > 0] &= E[v | X_2, u > -X_1 \beta_1] = E\left[\frac{\sigma_{uv}}{\sigma_u} u + \xi | u > -X_1 \beta_1\right] \\ &= \frac{\sigma_{uv}}{\sigma_u} E[u | u > -X_1 \beta_1] = \frac{\sigma_{uv}}{\sigma_u} \frac{\phi\left(-\frac{X_1 \beta_1}{\sigma_u}\right)}{1 - \Phi\left(-\frac{X_1 \beta_1}{\sigma_u}\right)} \\ &= \frac{\sigma_{uv}}{\sigma_u} \frac{\phi\left(-\frac{X_1 \beta_1}{\sigma_u}\right)}{\Phi\left(\frac{X_1 \beta_1}{\sigma_u}\right)} = \frac{\sigma_{uv}}{\sigma_u} \lambda\left(-\frac{X_1 \beta_1}{\sigma_u}\right) \end{aligned} \quad (14)$$

as it can be written that $v = \frac{\sigma_{uv}}{\sigma_u} u + \xi$ and $\lambda(z) = \frac{\phi(z)}{1 - \Phi(z)}$ is the Mills ratio where $\phi(\cdot)$ is probability density standard normal function and $\Phi(\cdot)$ is cumulative standard normal distribution function. The only missing part are the β_1 coefficients. Hence, Heckman proposed the estimation procedure as follows:

- estimate the probit model for the participation equation posing $\sigma_u = 1$ and obtain the $\hat{\beta}_1$,
- estimate values of $\hat{\lambda}$ for each observation,
- estimate the outcome equation by OLS on non-zero observations including $\hat{\lambda}$ as one of the regressors.

Heckman (1979) added that the usual formula for standard errors of coefficients in the OLS estimation is not valid. He derived further the appropriate formula in order to test the correlation of error terms by common t-statistics for $H_0 : \frac{\sigma_{uv}}{\sigma_u} = 0$.

All the previous estimators of models rely heavily on the distributional assumption of error terms as on its basis the MLE is built up. An alternative approach was presented by Cosslett (1991) who proposed a semi-parametric estimator of sample selection model instead of Heckman two-step estimator. The bias of the OLS estimator is described by the equation (13). Cosslett (1991) rewrote it in the form of a function

$$y_2 = X_2\beta_2 + \epsilon + \Psi(X_1\beta_1). \quad (15)$$

In general, concerning the participation equation the distributional assumption is not imposed and $F(-X_1\beta_1)$ of binary choice model is estimated by non-parametric technique. In the second stage, the functional form of $\Psi(z)$ in the outcome equation has to be specified. Few methods to obtain distribution $F(-X_1\beta_1)$ are discussed in the original paper, for example the one described in Cosslett (1983), or in Klein and Spady (1993) that is basically an implementation of kernel estimation. The consistent estimator of $F(-X_1\beta_1)$ is then used in the maximum likelihood function to obtain $\hat{\beta}_1$ coefficients. Afterwards, the function $\Psi(z)$ is specified by Cosslett (1991) as a set of dummy variables:

$$y_{2,i} = X_{2,i}\beta_2 + \epsilon_i + \sum_{j=1}^J \lambda_j I(i \in I(j)), \quad (16)$$

where $I(\cdot)$ is an indicator function and

$$I(j) = \{i | u_{j-1}^* < -X_{1,i}\hat{\beta}_1 < u_j^* \text{ and } y_{1,i} = 1\} \quad (17)$$

is the subset of observations with $u_0^* = -\infty$, $u_{J+1} = \infty$, $I(J+1)$ is empty, $I(1)$ can be empty, and remaining subsets are non-empty. The step function \hat{F} determines the location of steps u_j^* and their number J . As a complete set of dummy variables is used in the estimation, the constant term is omitted in order to avoid a dummy variable trap. To estimate the missing constant term, Cosslett (1991) proposes:

$$\hat{\beta}_{0,2} = \sum_{j=1}^M w_j \hat{\lambda}_j, \quad M = o(J), \quad (18)$$

where w_j are positive weights, closer unspecified. Finally, the outcome equation is estimated by OLS using the non-zero observations only.

To give an idea of employment of presented models and their estimators a few examples are given in the Table 1. The examples are taken from different areas of demand and policy analyses. The models are used for demand for energy resources (electricity, renewable sources), demand

Table 1: Summary of presented models and their estimators

Model	N°of processes	Errors (u,v)	Observations for estimation	Examples of application
Tobit	1	-	All	Zarnikau (2003) - Demand for renewable energy; Gifford and Bernard (2006) - Factors influencing demand for organic food
Double-hurdle	2	indep.	All	Newman et al. (2001) - Demand for prepared meals; Jones and Yen (2000) - Demand for beef
Two-part	2	indep.	Uncensored (non-zero)	Heres-Del-Valle and Niemeier (2011) - Factors influencing vehicle miles travelled; Deb and Trivedi (2002) - Demand for health care
Sample selection: FIML	2	corr.	All	Yen and Huang (2002) - Demand for beef
Sample selection: LIML - Heckman two-step	2	corr.	All in 1 st step, uncensored in the 2 nd step	Alviola and Capps (2010) - Demand for conventional and organic milk Carter et al. (2012) - Demand for electricity
Sample selection: semiparametric Cosslett	2	corr.	All in 1 st step, uncensored in the 2 nd step	Hussinger (2008) - Effect of public R&D subsidies

for different food products (beef, prepared meals), or analyses of factors influencing demand for organic food products.

Beside the application of different models and estimators, there is a discussion of their performance. We can find proponents of two-part model and of sample selection model in the literature. A few Monte Carlo studies were conducted to compare the approaches and their estimators. Most of them are cited and commented in Puhani (2000). The impact of different factors on the performance are examined such as the violation of distributional assumptions, the level of censoring, the level of correlation of disturbances, the collinearity between regressors, or the magnitude of variances of disturbances. The problem of studies are that they are not very extended. For example, the number of replications in Monte Carlo simulations is not high, the number of variables is usually two in both equations and sometimes they are simulated using uniform distribution.

Arabmazar and Schmidt (1982) have compared the censored and truncated tobit models when the distributional assumption of errors was violated and for different levels of censoring. For low levels (till 25 %) of censoring there is almost no significant bias and tobit model for truncated data records higher bias than the model for censored data. Manning et al. (1987) have compared the sample selection model with the two-part model and they conclude that the two-part model has better results than the other model also in case when the sample selection model was the true model. The article of Leung and Yu (1996) is the example of the paper that also summarizes some of the previous Monte Carlo studies. The authors criticised some of the discussed studies and in the own experiment they have found that without collinearity problems between regressors the

sample selection model performs better in case that it is true model and two-part model performs better when it is a true model but the sample selection model is close to its performance as well. Finally, Puhani (2000) summarizes main previous studies concluding that with higher correlation of errors the FIML performs better than LIML, the violation of assumption of joint distribution of errors does not pose significant problems and the most interesting is the problem of collinearity of regressors, with the serious collinearity the two-part model is more robust.

3 Methodology

In the literature review four models are presented in total, tobit model, double-hurdle model, two-part model and sample selection model. The last named is presented with three different possible estimators. Together with OLS estimator whose results are in case of censored data biased, there are seven presented ways how to evaluate the demand in single-equation framework.

The purpose of this paper beside the review of different models is to show the necessity of treatment methods and their different performance on the example of empirical application. All seven methods are applied on the data in order to estimate the demand for different items. As the elasticities are the main interest in the demand analysis, the log-log form of the models is estimated:

$$\text{participation equation: } y_1^* = X_1\beta_1 + u \quad (19)$$

$$\text{outcome equation: } \ln q^* = X_2\beta_2 + v \quad (20)$$

$$\text{observed variable: } \ln q = \begin{cases} X_2\beta_2 + v & y_1^* > 0 \\ T & \text{otherwise,} \end{cases} \quad (21)$$

where y_1^* is a dummy variable equal to 1 when $q > 0$, q is the quantity purchased, $y_2 = \ln q$, and T is the threshold value. Moreover, we are mainly interested in the outcome equation, which leads to the estimation of elasticities. The specific conditions of all methods are presented below.

TOBIT In the tobit model, only the outcome equation is considered. The log-likelihood function used for the estimation has following form (Tobin, 1958):

$$\begin{aligned} \ln L_N = & \sum_i d_i \cdot \left(-\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln \sigma_v^2 - \frac{1}{2\sigma_v^2} (y_{2,i} - x'_{2,i}\beta_2)^2 \right) \\ & + (1 - d_i) \cdot \ln \left(\Phi \left(\frac{T - x'_{2,i}\beta_2}{\sigma_v} \right) \right), \end{aligned} \quad (22)$$

where d_i is the dummy variable equal to 1 when $y_1^* > 0$.

HURDLE The double-hurdle model estimates β_1 and β_2 at the same time by the MLE upon the log-likelihood function (Jones, 1989):

$$\begin{aligned} L_N = & \prod_0 1 - \Phi \left(\frac{X_1\beta_1}{\sigma_u} \right) \Phi \left(\frac{X_2\beta_2 - T}{\sigma_v} \right) \prod_+ \Phi \left(\frac{X_1\beta_1}{\sigma_u} \right) \Phi \left(\frac{X_2\beta_2 - T}{\sigma_v} \right) \\ & \prod_+ \frac{1}{\sigma_v} \phi \left(\frac{y_2 - X_2\beta_2}{\sigma_v} \right) \Phi \left(\frac{X_2\beta_2 - T}{\sigma_v} \right)^{-1}, \end{aligned} \quad (23)$$

\prod_+ is the product over observations where $y_1^* > 0$ and $q > 0$ and \prod_0 is the product over the rest of observations.

TWO PART The outcome equation of two-part model can be estimated by OLS using the positive observations of quantity only.

FIML This estimator of sample selection model also estimates both β_1 and β_2 at the same time upon the log-likelihood function (Amemiya, 1984; Leung and Yu, 1996):

$$L_N = \prod_0 \Phi\left(\frac{-X_1\beta_1}{\sigma_u}\right) \prod_+ \Phi\left(\frac{\frac{X_1\beta_1}{\sigma_u} + \rho(y_2 - X_2\beta_2)/\sigma_v}{\sqrt{1-\rho^2}}\right) \phi\left(\frac{(y_2 - X_2\beta_2)}{\sigma_v}\right) \frac{1}{\sigma_v}, \quad (24)$$

where $\rho = \frac{\sigma_{uv}}{\sigma_u\sigma_v} = \frac{\sigma_{uv}}{\sigma_v}$ is the correlation coefficient of the error terms u and v and $\sigma_u = 1$ is assumed.

LIML We exactly follow the procedure described in Heckman (1979). In the first stage, we estimate β_1 by probit model, then we calculate $\hat{\lambda}$. In the second stage, we estimate β_2 by OLS using only uncensored (positive) observations and between the regressors X_2 we add the $\hat{\lambda}$. The log-likelihood function of the probit model we use is:

$$\ln L_N = \sum d \cdot \ln(1 - \Phi(-X_1\beta_1)) + (1 - d) \cdot \ln(\Phi(-X_1\beta_1)). \quad (25)$$

COSSLETT In the first stage, we estimate the $\hat{F}(X_1\hat{\beta}_1)$ upon log-likelihood function:

$$\ln L_N = \sum d \cdot \ln(1 - F(-X_1\beta_1)) + (1 - d) \cdot \ln(F(-X_1\beta_1)), \quad (26)$$

where the distribution function F is specified by Gallant and Nychka (1987). Their estimate of distribution F is close to one in Cosslett (1983) with the difference that it is a continuous estimate without need of special attention when maximizing the log-likelihood function. The approach of their estimation relies on assumption that the density function has the Hermite form

$$f(u) = [P_K(u)\phi(u)]^2, \quad (27)$$

where $P(u)$ is a polynomial of order K , K is increasing in the sample size and it is arbitrarily chosen. The overview of the estimator together with the description step by step can be found in Stewart (2004). Next, the outcome equation is estimated in the form of equation (16) by OLS using the non-zero observations only. To create the classes for dummy variables (borders u_j), the $X_1\hat{\beta}_1$ is split into intervals of equal length. Their number is determined in order the estimates are stable. The constant term has to be excluded from the equation as all dummy variables from the set $I(j)$ (equation 17) are used. In order to identify the constant term for the comparison between models, we follow the Hussinger (2008) approach. The constant term is equal to the dummy variable that represents the highest probability in participation equation. It means that the first dummy variable should be considered as a constant term as:

$$P[y_1 = 1|X_1] = 1 - P[u \leq -X_1\beta_1] = 1 - F(-X_1\beta_1). \quad (28)$$

OLS Finally, we would like to compare all the previous estimates with usual OLS estimate of outcome equation on all observations. As the quantity 0 cannot be used in the logarithm, the threshold value T is assigned to the observations where $q = 0$.

Together with estimates the standard errors are needed. In case of MLE, the usual formulas are valid only under the homoskedasticity of disturbances. The information matrix equality has to hold. In case of OLS estimates, Heckman (1979) argued that common formula for standard errors does not hold in Heckman two-step estimator. In order to overcome all these possible problems, we apply bootstrap method to estimate the standard errors in case of all 7 estimators. Specifically, we use paired bootstrap method that is able to capture the heteroskedasticity. The observations of y_1 , X_1 , y_2 , and X_2 are fixed and resampled with recurrence. For each newly created sample, the estimates are obtained in the usual way. In sum, we create $B = 500$ samples and new estimates $\hat{\beta}_b$ for each estimator. The formula for standard errors of original sample is then:

$$SE(\hat{\beta}_i) = \sqrt{\frac{1}{1 - B} \sum_{b=1}^B (\hat{\beta}_{i,b} - \hat{\beta}_i)^2}. \quad (29)$$

To evaluate the performance of models and estimators we use following criteria. For each analysed item, we have together 7 different estimates. Tobit model, double hurdle model, and FIML estimator are estimated upon the MLE method. We receive the value of log-likelihood. The rest, two-part model, LIML, Cosslett and OLS are estimated by OLS technique. The estimates of OLS are equal to estimates of MLE in Gaussian case, it means when the distribution of errors is normal. As for the estimates by MLE we assume normal distribution, we are able to obtain the log-likelihood value for the models and estimators estimated by OLS as well. We take into account the number of regressor as well as double hurdle model or FIML estimate more coefficients in the same time than in case of two-part model, for instance. For each estimate beside the value of log-likelihood we consider the value of Akaike information criterion and Schwarz Bayesian information criterion defined respectively according to Cameron and Trivedi (2005):

$$AIC = 2k - 2 \ln L \quad (30)$$

$$SBIC = \ln N \cdot k - 2 \ln L \quad (31)$$

where k is the number of estimated parameters by the MLE, $\ln L$ is the value of log-likelihood and N is the number of observations. The model with minimized criterion is preferred.

4 Data description

The dataset is compiled from the household-level budget survey that describes yearly expenditures and incomes for the year 2015 of 2,887 families living in the Czech Republic. The Czech Statistical Office collects every year the data from approximatively 3,000 households. This size represents roughly one thousandth of Czech households in total. The households are chosen on the basis of quotas that reflect the economic activity of the principal of the household as a primary aspect and the income of the household as a secondary aspect. The income classes included in the dataset do not exactly reflect the income composition of Czech households, thus the results

regarding the income are not representative in general. The dataset includes socio-demographic information about household's members, describes housing and equipment, registers all income and expenditure flows (monetary and barter as well), and records the quantity of purchased food items and in kind consumptions for some goods as well.

Although we use the yearly data, the expenditures for food items are collected in two separate months of the year and the data is then recalculated for the whole year. It means that the zeros are not generated only by the fact that the consumers are not interested in the given product but the infrequency of purchase is also a possible explanation. Hence, the censoring can really apply in this case.

The food products we have chosen for the analysis are listed in the Table 2 below. The first group of products are those with low level of censoring, we have chosen milk and poultry as products that consist of main nutrients and they have widespread usage on contrary with chocolate and chocolate products that are rather perceived as a dainty. The decision process of these two types of products can be realized differently. Same characteristics apply for pasta and mineral water with moderate level of censoring. In the group of products with medium level of censoring, there are potatoes, rice, and sugar. Potatoes and rice are there as products with similar use and representing the frequent component of main dishes. On the other hand, there is sugar that is usually used to sweeten hot drinks or for baking. The products with high level of censoring are beef and fish, these are products representing the important source of proteins but with higher price than poultry or pork. There are juices as well in this group as once again there exist cheaper alternatives of non-alcoholic drinks. In the last group of products with extreme level of censoring there are legumes, which have usually lower price and high nutrient value in contrast with cocoa that has higher price and it is not widespread used.

Table 2: Food products chosen for the analysis

Commodity	Level of censoring (%)
Milk	3.50
Poultry	6.86
Chocolate	9.32
Pasta	15.38
Mineral water	16.56
Sugar	23.45
Potatoes	24.94
Rice	30.59
Juices	43.44
Beef	46.83
Fish	47.49
Legumes	70.77
Cocoa	71.98

The variables used for the analysis are defined in the Table 3 and their descriptive statistics are given in the Table 4 and 5 below. On average, total household net income is 30,930 CZK

a month (about 1,200 Euro) and they spend 17 % for food, with 2 % and 58 % as the minimum and the maximum value, respectively. This share is a bit larger than the average share spent by Europeans that was 12.4 % in 2014 (OECD, 2016). There are 2.34 members in a household (1.86 units after weighting) with 52 years old head of family. In most families a female is present, about 29 % are households with a retired person and 40 % have at least one child. Quite large proportion of households can use in kind (or self-produced) food products, varying between 10 % (milk) and 43 % (potatoes), with only 2 % working in agricultural sector

Table 3: Description of variables

Variable	Description
<i>purchase_i</i>	dummy variable equal to 1 if the item <i>i</i> is purchased
<i>q_i</i>	yearly quantity of item <i>i</i> in kilograms or litres
<i>p_i</i>	price of item <i>i</i> defined as unit value (expenditures on <i>i</i> divided by <i>q_i</i>)
<i>income</i>	monthly income of the household in thousands of Czech crowns
<i>food_share</i>	expenditures on food in total relative to income (values between 0 and 1)
<i>size</i>	number of members living in the household
<i>units</i>	number of consumer units in the household; principal of the household is equivalent to 1, child between 0-13 years old is equivalent to 0.5; other members are equivalent to 0.7
<i>age</i>	age of the principal of the household that approximate the age category of the household
<i>woman</i>	dummy variable equal to 1 if there lives a woman in the household
<i>d_retired</i>	dummy variable equal to 1 if there lives a retired person in the household
<i>d_children</i>	dummy variable equal to 1 if there is at least one child in the household
<i>children_9</i>	number of children less than 9 years old
<i>primary</i>	dummy variable equal to 1 if the the highest achieved level of education in the households is primary school at maximum
<i>secondary</i>	dummy variable equal to 1 if the the highest achieved level of education in the household is on the secondary level (high school or equivalent)
<i>tertiary</i>	dummy variable equal to 1 if the the highest achieved level of education in the household is higher than high school
<i>natu_i</i>	dummy variable equal to 1 if the household obtains given item <i>i</i> from the kept animals or as a gift (i.e., in kind consumption if available)
<i>unemployed</i>	the number of unemployed members of the household
<i>agriculture</i>	dummy variable equal to 1 if there is a member of the household working in agriculture, forestry, fishing, hunting
<i>nonurban</i>	dummy variable equal to 1 if the household is located in the sparsely populated area
<i>village</i>	dummy variable equal to 1 if the household lives in the village
<i>freezer</i>	dummy variable equal to 1 if the household owns a freezer to store food products

Table 4: Descriptive statistics of variables

Variable	Mean	Min	Max	Std
<i>income</i>	30.93	3.26	234.62	16.96
<i>food_share</i>	0.17	0.02	0.58	0.07
<i>size</i>	2.34	1.00	7.00	1.16
<i>units</i>	1.86	1.00	4.80	0.72
<i>age</i>	51.82	20.00	90.00	14.68
<i>woman</i>	0.93	0.00	1.00	0.26
<i>d_retired</i>	0.29	0.00	1.00	0.45
<i>d_children</i>	0.40	0.00	1.00	0.49
<i>children_9</i>	0.27	0.00	3.17*	0.60
<i>primary</i>	0.04	0.00	1.00	0.18
<i>secondary</i>	0.76	0.00	1.00	0.43
<i>tertiary</i>	0.21	0.00	1.00	0.40
<i>natu_milk</i>	0.10	0.00	1.00	0.30
<i>natu_poultry</i>	0.20	0.00	1.00	0.40
<i>natu_potatoes</i>	0.43	0.00	1.00	0.50
<i>natu_beef</i>	0.12	0.00	1.00	0.32
<i>unemployed</i>	0.08	0.00	3.33*	0.28
<i>agriculture</i>	0.02	0.00	1.00	0.14
<i>nonurban</i>	0.33	0.00	1.00	0.47
<i>village</i>	0.33	0.00	1.00	0.47
<i>freezer</i>	0.92	0.00	1.00	0.27

*Decimal numbers stands for number of months
in given category

5 Model specification

For the models with decision making in two steps we have to select variables for each step separately. For the first step, we have been considering variables with following motivations. The size of household, i.e. the number of its members, gives higher probability that the product will be demanded by some of its members. The age category of household can reflect the change in preferences or taste with age. The dummy variable indicating the presence of adult woman in the household can incorporate specific female demands. The same holds for dummy variables indicating the presence of children or retired person in the household. We also distinguish different levels of education of the person with the highest education in the household. It is possible that more educated inhabitants care more about the products they eat and thus prefer different food products than others. Next, we control also for location of households. We consider the locations with very low density of population where the supply and availability of products is different than in very urbanized areas and the prices can be higher. Then, we consider also the location in villages where beside the availability of products the diet habits can be different and the barter

Table 5: Descriptive statistics for prices and quantities for non-zero observations

Food	Var	Mean	Min	Max	Std	Var	Mean	Min	Max	Std
Milk	p_1	13.31	8.00	37.05	3.55	q_1	41.24	0.81	270.49	35.75
Poultry	p_2	83.35	15.60	235.19	27.44	q_2	13.04	0.24	84.57	10.43
Chocolate	p_3	222.87	59.77	877.78	87.43	q_3	2.32	0.05	19.93	2.17
Pasta	p_4	42.38	8.50	211.25	20.91	q_4	4.75	0.12	37.96	4.15
Water	p_5	8.63	1.50	50.00	4.78	q_5	58.82	0.57	565.54	65.15
Sugar	p_6	18.17	7.33	115.97	10.53	q_6	13.00	0.33	138.84	14.29
Potatoes	p_7	11.91	1.11	33.51	3.77	q_7	27.77	0.58	438.75	38.59
Rice	p_8	45.42	14.00	301.43	30.36	q_8	3.69	0.11	37.38	3.50
Juices	p_9	25.55	7.00	115.05	9.59	q_9	9.93	0.28	105.95	10.96
Beef	p_{10}	155.96	47.56	398.89	42.96	q_{10}	4.30	0.13	41.77	4.25
Fish	p_{11}	172.20	50.00	495.13	79.02	q_{11}	3.41	0.16	28.05	3.45
Legumes	p_{12}	49.55	12.99	267.83	24.62	q_{12}	1.84	0.10	10.00	1.47
Cocoa	p_{13}	173.88	50.00	441.43	63.46	q_{13}	0.92	0.08	7.82	0.89

flows of food can be more frequent. Another factor linked to purchase of certain items can be an unemployment. It is connected with lower income and in such period the more expensive products can be excluded from the diet or substituted, e.g. beef vs poultry meat. Finally, we consider the workers in agriculture can have different approach to some products and we consider also the other ways to obtain given product than purchase (gifts or gardening, it means in kind consumption).

Concerning the second step, the outcome equation, we consider mostly the quantitative factors in contrast with participation equation where the qualitative factors are mainly used. Beside the prices and income the expenditures on all food goods relative to income are considered as the food budget share may be correlated with the household expenditures on the given food product and its price. Then, the consumer units can influence the quantity purchased as more consumers need more food. The presence of small children can beside the composition of purchased products determine its quantity. We also incorporate in the model the age category of the household to control different needs with respect to age. The location of household in the village can beside the higher prices and less frequent availability of products be linked with natural consumption of given products and thus the quantity needed to be purchased would be smaller than for the other households. Finally, we consider the presence of freezer in the household that enables to store bigger quantity of given products. In our case, it means poultry, beef and fish. Hence, the participation equation has the following form:

$$\ln q_i = \beta_{2,0} + \beta_{2,1} \ln p_i + \beta_{2,2} \ln income + \beta_{2,3} food_share + \beta_{2,4} units + \beta_{2,5} age + \beta_{2,6} children_9 + \beta_{2,7} village + (\beta_{2,8} freezer) + v. \quad (32)$$

We assume the double-log specification to estimate the outcome equation through which it is straightforward to get the price and income elasticities. Since $\ln q_i$ is not defined for zero consumption, we impute a threshold value $T = -3$, hence for $q_i \leq 0$ the value is $\ln q_i = T$. The threshold value is chosen as the value lower than minimal value of observed non-zero consumption in the

dataset. The exact value of threshold does not have an impact on the demand estimation. Regarding the prices, these are defined as unit values. For zero observations the unit values are not available. Therefore, the sample average value is assigned to missing prices such as in Yen et al. (2002).

A special case is the tobit model. It considers variables that influence the participation and outcome results in the same time. Hence, for the tobit model we consider a mixture of variables previously described for participation and outcome equation.

6 Results and Discussion

Firstly, we conduct preliminary estimates of participation equation via probit model using the variables described in the previous section. The results given by the final choice of variables are displayed in the Appendix Tables A.1-A.2. For each food product different composition of variables is used according to specific features of the product and significance of the linked variables. The variables given in the named tables are used in the participation equations of the other estimators that are not specified by probit model.

Next, the complete results of outcome equation (equation (32)) for all estimators are displayed in the Appendix Tables A.3.-A.15. Concerning the tobit model, all the results for the mixture of variables are displayed there. Regarding the results of other models and their estimators, only the results of the outcome equation are given as the results of participation equation are not of interest. For example, in case of double-hurdle model and FIML estimator the participation and outcome equations are estimated in one step by MLE but the results of $\hat{\beta}_2$ are presented only. In case of Cosslett's semi-parametric estimator the results of coefficients at dummy variables are omitted as well.

The estimates of elasticities are the most important in the demand analyses and the results of comparison criteria are of the interest in the empirical part of the paper. The elasticities' results are given in the Table 6 and 7. The preferred models or estimators according to criteria are listed in the Table 8.

To remind, for all the examined models and estimators the coefficients at $\ln p_i$ and $\ln income$ are directly equal to the value of elasticities except the two-part model. This model assumes that the participation equation is modelled conditionally (equation 12) but we are interested in the elasticities for whole population instead of elasticities for the non-zero subsample. The estimate of elasticities is then given as follows:

$$E[\ln(y_2)] = P[y_1^* > 0] E[\ln(y_2|y_1^* > 0)] = \Phi(X_1\beta_1) X_2\beta_2, \quad (33)$$

$$\frac{\partial E[\ln(y_2)]}{\partial x_k} \Big|_{x_k \in X_2, x_k \notin X_1} = \Phi(X_1\beta_1) \beta_{2,k}. \quad (34)$$

Hence, the elasticities for two-part model are evaluated according to equation (34) at sample mean of regressors \bar{X}_1 .

Comparing the values of elasticities among different estimators we can directly observe that values produced by tobit model and by OLS are visibly different from values of other estimators. Moreover, the values of log-likelihood, of *AIC*, and *SBIC* are more distant from the others as well. With higher level of censoring, the differences between these two estimators and the rest of them are bigger. It is also remarkable that even for low levels of censoring (7% for poultry) the

Table 6: Estimates of own-price elasticities

Product	Own-price elasticities						
	TOBIT	HURDLE	TWO PART	FIML	LIML	COSS	OLS
Milk	-1.2031***	-1.031***	-1.0792 _I ***	-0.7463***	-1.1031***	-1.1006 _L ***	-1.253***
Poultry	-0.9049***	-0.4870***	-0.5789***	-0.0513	-0.6091***	-0.6030 _{L,I} ***	-0.8850***
Chocolate	-0.8172***	-0.1934***	-0.4518 _I ***	-0.2399***	-0.4942***	-0.4948 _L ***	-0.7559***
Pasta	-1.0505***	-0.3956***	-0.4120 _I ***	-0.4870***	-0.4779***	-0.4749 _L ***	-0.8827***
Water	-2.3659***	-1.1901***	-0.9981***	-1.1878***	-1.1853 _{L,I} ***	-1.1835***	-1.7632***
Sugar	-2.6440***	-1.0299***	-0.8091***	-0.9527***	-1.0377 _{L,I} ***	-1.0410***	-1.8189***
Potatoes	-2.7336***	-1.0076***	-0.8783***	-0.9608***	-1.1168 _I ***	-1.1143 _L ***	-1.9993***
Rice	-2.5805***	-0.5495***	-0.5319***	-0.7821***	-0.7639 _I ***	-0.7585 _L ***	-1.6325***
Juices	-5.1051***	-0.6680***	-0.4376***	-0.4783***	-0.7791 _{L,I} ***	-0.7702***	-2.0642***
Beef	-3.5663***	0.1081***	0.0574	-0.0159	0.1065	0.1061 _{L,A}	-1.1592***
Fish	-4.2100***	-0.4234***	-0.3921 _I ***	-0.4262***	-0.7448***	-0.7423 _L ***	-1.9166***
Legumes	-6.5334***	-0.1680***	-0.1161 _I ***	-0.3747	-0.4045***	-0.4160 _{L,A} ***	-2.0631***
Cocoa	-3.7465***	-0.2934***	-0.3337***	-0.8225***	-1.2198***	-1.2146 _{L,I} ***	-2.4392***

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

L stand for preferred model according to the value of $\log L$, I for both information criteria, A for AIC

Table 7: Estimates of income elasticities

Product	Income elasticities						
	TOBIT	HURDLE	TWO PART	FIML	LIML	COSS	OLS
Milk	0.7989***	0.4543***	0.4838 _I ***	0.4452***	0.4925***	0.4940 _L ***	0.7281***
Poultry	1.4643***	0.4912***	0.7069***	0.4224***	0.7455***	0.7469 _{L,I} ***	1.2636***
Chocolate	1.3700***	0.3693***	0.7405 _I ***	0.3786***	0.7918***	0.7845 _L ***	1.3261***
Pasta	1.2079***	0.1742***	0.3864 _I ***	0.2897***	0.4498***	0.4268 _L ***	0.9440***
Water	2.8788***	0.4510***	0.6043***	0.5351***	0.6521 _{L,I} ***	0.6927***	1.6443***
Sugar	1.4709***	0.2788***	0.3066***	0.3003***	0.4208 _{L,I} ***	0.4173***	0.7190***
Potatoes	1.9762***	0.3508***	0.4247***	0.4611***	0.5455 _I ***	0.5447 _L ***	1.1251***
Rice	1.5015***	0.2554***	0.3140***	0.2744***	0.4589 _I ***	0.4662 _L ***	0.8448***
Juices	5.0319***	0.6896***	0.4604***	0.7268***	0.7704 _{L,I} ***	0.7874***	1.8774***
Beef	7.3006***	0.2869***	0.3345***	0.3848	0.6417***	0.6424 _{L,A} ***	2.4251***
Fish	3.0483***	0.4144***	0.3102 _I ***	0.4964***	0.6044***	0.6217 _L ***	1.2682***
Legumes	2.7136***	0.1973***	0.0692 _I ***	0.0859	0.2416***	0.2342 _{L,A} ***	0.6284***
Cocoa	1.005***	0.3037***	0.0984***	0.4171***	0.3489***	0.3967 _{L,I} ***	0.3307***

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

L stand for preferred model according to the value of $\log L$, I for both information criteria, A for AIC

results of OLS and non-zero subsample OLS (two-part model) are already different. The Heckman two-step estimator and Cosslett's semi-parametric estimator of sample selection model have very close values to each other for all food products. The same holds for the comparison criteria. The two-part model is close to Heckman two-step estimator and Cosslett's estimator for products with low or medium level of censoring, with higher censoring (50 % and more) the results become to be significantly different. On contrary, the values of comparison criteria are close to those two estimators even with high level of censoring. Estimates produced by FIML are rather variable. For low and moderate level of censoring, the values are close to those of Heckman two-step estimator and Cosslett's estimator with some exceptions (own-price elasticity for poultry or income elasticity for pasta). With increasing levels of censoring, the estimates are more distant from LIML and Cosslett but not as distant as the two-part results. The comparison criteria remains worse than for two-part model, LIML and Cosslett's estimator. It is partly linked to the fact that the FIML together with double-hurdle model estimate more parameters in the MLE routine than cited ones as they give the estimates of coefficients of participation and outcome equation in the same time. Although both information criteria accounts for the number of parameters estimated by MLE routine, the differences in the log-likelihood value are still preserved in the information criteria values. On the other hand, the Cosslett's estimator gives also high number of coefficients due to the set of dummy variables in the second step. The double-hurdle model gives varying estimates as well. For some products the values are very close to those of LIML and Cosslett, for other products the values are close to FIML but more distant from LIML and Cosslett and some values are distant from all other estimators. There is no clear link to the level of censoring. The values of comparison criteria are similar to those of FIML estimator.

Table 8: List of estimators preferred according to different criteria

Product	Censoring	logL	AIC	SBIC
Milk	3.50	COSSLETT	TWO PART	TWO PART
Poultry	6.86	COSSLETT	COSSLETT	COSSLETT
Chocolate	9.32	COSSLETT	TWO PART	TWO PART
Pasta	15.38	COSSLETT	TWO PART	TWO PART
Water	16.56	LIML	LIML	LIML
Sugar	23.45	LIML	LIML	LIML
Potatoes	24.94	COSSLETT	LIML	LIML
Rice	30.59	COSSLETT	LIML	LIML
Juices	43.44	LIML	LIML	LIML
Beef	46.83	COSSLETT	COSSLETT	TWO PART
Fish	47.49	COSSLETT	TWO PART	TWO PART
Legumes	70.77	COSSLETT	COSS/2PART	TWO PART
Cocoa	71.98	COSSLETT	COSSLETT	COSSLETT

The Table 8 gives an overview of models preferred by various criteria. We can observe that the type of product as described in the section 4 can have little influence on the performance of estimators. For example, for medium level of censoring we have three types of goods - sugar, potatoes and rice. We expect similar motivation behind the demand for potatoes and rice in contrast with sugar and the performance of estimators in these cases slightly differs. It is also evident

that the performance of Heckman two-step estimator (LIML), Cosslett's semi-parametric estimator and two-part model are significantly better given the criteria we have preliminary specified. Moreover, there is no clear evolution of their performance with the level of censoring. The main drawback of these results is that the coefficients of Cosslett's estimator and Heckman two-step estimator are very close to each other in contrast to the results produced by two-part model.

In the Table 9, there are the absolute values of deviations of elasticities' estimates in percentage from the estimate of best performing estimator according to information criteria and according to value of log-likelihood for own-price elasticities. The Table 10 brings the values for income elasticities. The behaviour of estimators together with differences in their estimates are already described above, however through the deviations it is possible to quantify them. As the OLS generates by definition biased results its estimates of elasticities deviate enormously from best performing models. The same behaviour is followed by tobit model that supposes only one step decision process and as supposed it is not likely to describe the process correctly.

Among the models that assume two stochastic processes, double-hurdle is never the best performing model, its elasticities' estimates deviate from the elasticity values of best performing model up to more than 100 % but on average, considering all 13 items, the deviation is about 26 % for own-price elasticities and about 38 %. FIML estimator of the sample selection model is never the best estimator either. Deviations in the magnitude of its elasticities are up to 223 % (42 % on average across all items) for the own price elasticities, and up to 60 % (28 % on average) for income elasticities.

Considering AIC/BIC criteria, two-part model and LIML are the best performing for five food items each, and semi-parametric Cosslett is the best model for remaining three. However, there is no general recommendation for one best performing model. If we compare the elasticity estimates from the four well performing models and estimators (two-part, FIML, LIML, Cosslett), the deviation from the best performing is still quite large for each of them. If we consider the cases when particular model is not best performing, in the case of price elasticity, average deviation from the best performing is 30 % for two-part, 42 % for Cosslett, 57 % for LIML, and 49 % for FIML. In the case of income elasticities, the deviations are 30 %, 42 %, 59 %, and 26 %, respectively. From the point of view of log-likelihood value, basically Cosslett's estimator performs the best. And as the estimated value by Cosslett's estimator and by LIML estimator are close each other, the deviations from best performing model between these two estimators are almost negligible.

7 Conclusion

This paper deals with the problem of censored data in the consumer demand analysis conducted on household budget survey data. The censoring is caused by non-negligible portion of zero consumption recorded in the data and can lead to the selectivity problem. Untreated estimation procedures then give biased results. In the literature different methods are presented in order to work with censored data properly. In this paper, we summarize the tobit model, two-part model, double-hurdle model and sample selection model. The last named is presented with three different estimators - FIML (MLE routine), Heckman two-step procedure and Cosslett's semi-parametric technique. Then, we show that the treatment is important even for datasets with small portion of zeros and that the choice of the treatment method matters. Together with common OLS, the behavior of these different approaches to demand analysis is illustrated on the empirical example.

Table 9: Deviations of estimates of own-price elasticities from the estimates produced by best performing estimator according to information criteria and according to log-likelihood value

Deviations in percentage according to IC							
Product	TOBIT	HURDLE	TWO PART	FIML	LIML	COSS	OLS
Milk	11.48	4.47	best	30.85	2.21	1.98	16.10
Poultry	50.07	19.24	4.00	91.49	1.01	best	46.77
Chocolate	80.88	57.19	best	46.90	9.38	9.52	67.31
Pasta	154.98	3.98	best	18.20	16.00	15.27	114.25
Water	99.60	0.40	15.79	0.21	best	0.15	48.76
Sugar	154.79	0.75	22.03	8.19	best	0.32	75.28
Potatoes	144.77	9.78	21.36	13.97	best	0.22	79.02
Rice	237.81	28.07	30.37	2.38	best	0.71	113.71
Juices	555.26	14.26	43.83	38.61	best	1.14	164.95
Beef	6313.07	88.33	best	127.70	85.54	84.84	2119.51
Fish	973.71	7.98	best	8.70	89.95	89.31	388.80
Legumes	5527.39	44.70	best	222.74	248.41	258.31	1677.00
Cocoa	208.46	75.84	72.53	32.28	0.43	best	100.82
AVERAGE	1116.33	27.31	29.99	49.40	56.62	41.98	385.56
Deviations in percentage according to log-likelihood							
Milk	9.31	6.32	1.94	32.19	0.23	best	13.85
Poultry	50.07	19.24	4.00	91.49	1.01	best	46.77
Chocolate	65.16	60.91	8.69	51.52	0.12	best	52.77
Pasta	121.20	16.70	13.24	2.55	0.63	best	85.87
Water	99.60	0.40	15.79	0.21	best	0.15	48.76
Sugar	154.79	0.75	22.03	8.19	best	0.32	75.28
Potatoes	145.32	9.58	21.18	13.78	0.22	best	79.42
Rice	240.21	27.55	29.87	3.11	0.71	best	115.23
Juices	555.26	14.26	43.83	38.61	best	1.14	164.95
Beef	3461.26	1.89	45.90	114.99	0.38	best	1192.55
Fish	467.16	42.96	47.18	42.58	0.34	best	158.20
Legumes	1470.53	59.62	72.09	9.93	2.76	best	395.94
Cocoa	208.46	75.84	72.53	32.28	0.43	best	100.82
AVERAGE	542.18	25.85	30.64	33.96	0.69	0.52	194.65

Table 10: Deviations of estimates of income elasticities from the estimates produced by best performing estimator according to information criteria and according to log-likelihood value

Deviations in percentage according to IC							
Product	TOBIT	HURDLE	TWO PART	FIML	LIML	COSS	OLS
Milk	65.13	6.10	best	7.98	1.80	2.11	50.50
Poultry	96.05	34.23	5.36	43.45	0.19	best	69.18
Chocolate	85.01	50.13	best	48.87	6.93	5.94	79.08
Pasta	212.60	54.92	best	25.03	16.41	10.46	144.31
Water	341.47	30.84	7.33	17.94	best	6.23	152.15
Sugar	249.55	33.75	27.14	28.64	best	0.83	70.87
Potatoes	262.27	35.69	22.14	15.47	best	0.15	106.25
Rice	227.20	44.35	31.58	40.20	best	1.59	84.09
Juices	553.15	10.49	40.24	5.66	best	2.21	143.69
Beef	2082.54	14.23	best	15.04	91.84	92.05	624.99
Fish	882.69	33.59	best	60.03	94.84	100.42	308.83
Legumes	3821.39	185.12	best	24.13	249.13	238.44	808.09
Cocoa	153.34	23.44	75.20	5.14	12.05	best	16.64
AVERAGE	694.80	42.84	29.86	25.97	59.15	41.86	204.51
Deviations in percentage according to log-likelihood							
Milk	61.72	8.04	2.06	9.88	0.30	best	47.39
Poultry	96.05	34.23	5.36	43.45	0.19	best	69.18
Chocolate	74.63	52.93	5.61	51.74	0.93	best	69.04
Pasta	183.01	59.18	9.47	32.12	5.39	best	121.18
Water	341.47	30.84	7.33	17.94	best	6.23	152.15
Sugar	249.55	33.75	27.14	28.64	best	0.83	70.87
Potatoes	262.81	35.60	22.03	15.35	0.15	best	106.55
Rice	222.07	45.22	32.65	41.14	1.57	best	81.21
Juices	553.15	10.49	40.24	5.66	best	2.21	143.69
Beef	1036.46	55.34	47.93	40.10	0.11	best	277.51
Fish	390.32	33.34	50.10	20.15	2.78	best	103.99
Legumes	1058.67	15.76	70.45	63.32	3.16	best	168.32
Cocoa	153.34	23.44	75.20	5.14	12.05	best	16.64
AVERAGE	360.25	33.70	30.43	28.82	2.67	3.08	109.82

The demand for different food products is estimated by all these methods. Food products are chosen because of great variety of level of censoring that is usually considered as one of the factors influencing the performance of estimators and it is easily observable factor.

By definition, OLS results in biased estimates and treatment method is needed. Tobit model that assumes the decision process being made in one step, i.e. there is only one stochastic process behind the decision about the purchase and quantity purchased, is always worst performing model with elasticity estimates relatively close to OLS. In our case, double-hurdle and FIML estimator of sample selection model are never best, but their elasticity estimates are on average close to the magnitudes of elasticities of the best performing estimators. Depending on the food item, two-part, LIML, and semi-parametric Cosslett estimator are the best performing. Without correction, own price elasticities may deviate from the ones estimated by the most suitable model significantly; up to 31 % for milk with 3.5 % of zeros and up to 60 % in case of income elasticity.

Analysing cross-section data for the year 2015, own price elasticities derived from all well performed treatment models are significant and negative for all analysed food items, but beef. In the case of beef it seems that households are not on average price sensitive. On average, Czech households respond to price changes relatively weakly for legumes, chocolate, pasta, poultry and fish for that the own price elasticity is around -0.4 to -0.6. Slightly large elasticity is estimated for rice and juices (-0.8), whereas households are quite sensitive towards changes in prices of milk, water, sugar, potatoes, and cocoa (-1.0 to -1.2).

The lowest income elasticity is for legumes and fish (0.7 and 0.31), whereas on the other hand there are juices, chocolate, and poultry with income elasticities around 0.77, followed by water and beef (0.65). The income elasticities for remaining food goods are between +0.40 and +0.55.

Our review of treatment methods for selectivity problems in single-equation demand analysis aims to summarize available methods and describe differences between them. Then, we want to point out that the treatment methods are important even when the small portion of zeros is contained in the dataset and that the choice has to be made carefully as the suitable model has to be employed in each case. In the demand analysis, it is very frequent that treatment methods are applied without broad discussion or careful selection. According to our empirical example, it seems that LIML and Cosslett estimator might be used when the share of zeros is relatively small (up to 10 %), but the two-part model is performing comparably to these two estimators of the sample selection model. If the share of zeros is larger, sample selection model should be used, and although we find that LIML is best performing among the estimators, when the chosen model is not performing the best the smallest deviation from the elasticity value is generated by Cosslett's estimator. If zeros exceed positive observations, two-part model or Cosslett's estimator seem to perform better but two-part model generates smaller deviations if it is not truly performing best.

References

- Alviola, P. A. and O. Capps (2010). Household demand analysis of organic and conventional fluid milk in the united states based on the 2004 nielsen homescan panel. *Agribusiness* 26(3), 369–388.
- Amemiya, T. (1974). Multivariate regression and simultaneous equation models when the dependent variables are truncated normal. *Econometrica* 42(6), pp. 999–1012.
- Amemiya, T. (1984). Tobit models: A survey. *Journal of Econometrics* 24(1–2), 3 – 61.
- Arabmazar, A. and P. Schmidt (1982). An investigation of the robustness of the tobit estimator to non-normality. *Econometrica* 50(4), pp. 1055–1063.
- Cameron, A. C. and P. K. Trivedi (2005). *Microeconometrics: methods and applications*. Cambridge university press.
- Carter, A., R. Craigwell, and W. Moore (2012). Price reform and household demand for electricity. *Journal of Policy Modeling* 34(2), 242–252.
- Cosslett, S. R. (1983). Distribution-free maximum likelihood estimator of the binary choice model. *Econometrica* 51(3), 765–82.
- Cosslett, S. R. (1991). Semiparametric estimation of a regression model with sample selectivity. *Nonparametric and semiparametric methods in econometrics and statistics*, 175–97.
- Cragg, J. G. (1971). Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica* 39(5), pp. 829–844.
- Deb, P. and P. K. Trivedi (2002). The structure of demand for health care: latent class versus two-part models. *Journal of health economics* 21(4), 601–625.
- Gallant, A. R. and D. W. Nychka (1987). Semi-nonparametric maximum likelihood estimation. *Econometrica: Journal of the Econometric Society*, 363–390.
- Gifford, K. and J. C. Bernard (2006). Influencing consumer purchase likelihood of organic food. *International Journal of Consumer Studies* 30(2), 155–163.
- Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. In *Annals of Economic and Social Measurement, Volume 5, number 4*, pp. 475–492. NBER.
- Heckman, J. J. (1979, January). Sample Selection Bias as a Specification Error. *Econometrica* 47(1), 153–162.
- Heres-Del-Valle, D. and D. Niemeier (2011). Co2 emissions: are land-use changes enough for california to reduce vmt? specification of a two-part model with instrumental variables. *Transportation Research Part B: Methodological* 45(1), 150–161.
- Hussinger, K. (2008). R&d and subsidies at the firm level: an application of parametric and semi-parametric two-step selection models. *Journal of Applied Econometrics* 23(6), 729–747.

- Jones, A. M. (1989). A double-hurdle model of cigarette consumption. *Journal of Applied Econometrics* 4(1), pp. 23–39.
- Jones, A. M. and S. T. Yen (2000). A box-cox double-hurdle model. *The Manchester School* 68(2), 203–221.
- Klein, R. W. and R. H. Spady (1993). An efficient semiparametric estimator for binary response models. *Econometrica* 61(2), pp. 387–421.
- Leung, S. F. and S. Yu (1996). On the choice between sample selection and two-part models. *Journal of Econometrics* 72(1–2), 197 – 229.
- Manning, W., N. Duan, and W. Rogers (1987). Monte carlo evidence on the choice between sample selection and two-part models. *Journal of Econometrics* 35(1), 59 – 82.
- Newman, C. F., M. Henchion, and A. Matthews (2001). A double-hurdle model of irish household expenditure on prepared meals. Economic papers, Trinity College Dublin, Economics Department.
- OECD (2016). Final consumption expenditure of households. [*electronic, last checked 01-16-2017*].
- Powell, J. L. (1986). Symmetrically trimmed least squares estimation for tobit models. *Econometrica* 54(6), pp. 1435–1460.
- Puhani, P. (2000). The heckman correction for sample selection and its critique. *Journal of Economic Surveys* 14(1), 53–68.
- Stewart, M. B. (2004). Semi-nonparametric estimation of extended ordered probit models. *The Stata Journal* 4(1), 27–39.
- Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica* 26(1), pp. 24–36.
- Yen, S. T. and C. L. Huang (2002). Cross-sectional estimation of us demand for beef products: a censored system approach. *Journal of Agricultural and Resource Economics*, 320–334.
- Yen, S. T., K. Kan, and S.-J. Su (2002). Household demand for fats and oils: two-step estimation of a censored demand system. *Applied Economics* 34(14), 1799–1806.
- Zarnikau, J. (2003). Consumer demand for ‘green power’ and energy efficiency. *Energy Policy* 31(15), 1661–1672.

Appendix

Table A.1: Results of probit model

	Milk	Poultry	Chocolate	Pasta	Water	Sugar
<i>constant</i>	1.298*** -0.229	0.238 -0.181	0.282 -0.184	0.473** -0.17	0.233 -0.154	-0.211 -0.153
<i>size</i>	0.260*** -0.0762	0.279*** -0.0592	0.238*** -0.0512	0.306*** -0.0469	0.243*** -0.042	0.137*** -0.0262
<i>age</i>	-0.00212 -0.00336	0.00369 -0.00272	0.0032 -0.00317	-0.00643* -0.00286	0.00442* -0.00221	0.00815** -0.0025
<i>woman</i>	0.359* -0.142	0.691*** -0.112	0.451*** -0.11	0.375*** -0.103	0.0389 -0.107	0.243* -0.0993
<i>d_children</i>	-0.159 -0.171	-0.048 -0.137	-0.0965 -0.119	-0.139 -0.109	-0.17 -0.0996	
<i>d_retired</i>			0.114 -0.104	-0.0998 -0.0873		0.0425 -0.0801
<i>primary</i>	0.186 -0.265	0.823* -0.329	-0.457** -0.149	0.00932 -0.148	-0.152 -0.142	0.0637 -0.146
<i>tertiary</i>	-0.313** -0.109	-0.245** -0.0914	0.0803 -0.09	-0.114 -0.0754	-0.0962 -0.0714	-0.273*** -0.0637
<i>nonurban</i>		-0.119 -0.0824			0.127* -0.0623	
<i>village</i>	-0.183 -0.0985					0.0611 -0.0573
<i>unemployed</i>		-0.145 -0.141	-0.330** -0.107		-0.214* -0.099	
<i>agriculture</i>						0.294 -0.222
<i>natu_milk</i>	-0.414** -0.129					
<i>natu_poultry</i>		-0.131 -0.0946				
logL	-411.1	-653.0	-842.4	-1153.7	-1260.2	-1530.9
pseudo R^2	0.0611	0.0952	0.0582	0.069	0.0274	0.0264

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.2: Results of probit model, con't

	Potatoes	Rice	Juices	Beef	Fish	Legumes	Cocoa
<i>constant</i>	0.594***	0.0881	0.194	-0.935***	-0.529***	-1.375***	-1.350***
	-0.161	-0.14	-0.147	-0.149	-0.099	-0.165	-0.156
<i>size</i>	0.0267	0.204***	0.175***	0.218***	0.194***	0.179***	0.252***
	-0.027	-0.0368	-0.0244	-0.0344	-0.0345	-0.0249	-0.0356
<i>age</i>	0.00648*	-0.00371	-0.00927***	0.00403		0.0023	0.0029
	-0.00264	-0.00197	-0.00236	-0.0024		-0.00251	-0.00212
<i>woman</i>	0.322**	0.235*	0.169	0.486***	0.220*	0.293*	0.0493
	-0.105	-0.0966	-0.0979	-0.0999	-0.0962	-0.115	-0.111
<i>d_children</i>		-0.088		-0.283***	-0.14		-0.0681
		-0.0887		-0.0853	-0.0827		-0.0889
<i>d_retired</i>	-0.181*		-0.0971	0.121	0.0317	0.0855	
	-0.084		-0.0726	-0.0735	-0.0587	-0.0764	
<i>primary</i>	0.219	-0.0288	-0.350*	-0.390**	-0.129	-0.163	-0.00652
	-0.167	-0.132	-0.137	-0.134	-0.131	-0.146	-0.142
<i>tertiary</i>	-0.0411	-0.0685	0.117	0.00322	0.0751	0.0658	-0.167**
	-0.0676	-0.0631	-0.0618	-0.0599	-0.0599	-0.0622	-0.0646
<i>nonurban</i>		-0.0727	-0.230***		-0.0498	-0.105	0.0675
		-0.0538	-0.0516		-0.0512	-0.0544	-0.0536
<i>village</i>	-0.455***			-0.127*			
	-0.0571			-0.0512			
<i>unemployed</i>	0.429***		-0.172*	-0.139	-0.188*		
	-0.114		-0.0867	-0.0876	-0.0866		
<i>agriculture</i>	-0.352	0.302			0.196	0.156	
	-0.182	-0.206			-0.178	-0.183	
<i>natu_potatoes</i>	-0.821***						
	-0.0547						
<i>natu_beef</i>				-0.108			
				-0.0731			
logL	-1428.4	-1727.2	-1869.5	-1932.7	-1957.0	-1700.2	-1655.1
pseudo R^2	0.1191	0.0284	0.054	0.0314	0.0203	0.0254	0.0335

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.3: Milk

	TOBIT	HURDLE	TWO PART	FIML	LIML	COSS	OLS
<i>constant</i>	3.1682*** (0.4567)	4.4412*** (0.2872)	4.4806*** (0.2256)	4.0461*** (0.5202)	4.5683*** (0.2401)	4.3822*** (0.2627)	3.5504*** (0.3619)
<i>ln p</i>	-1.2031*** (0.0754)	-1.031*** (0.1007)	-1.1105*** (0.0641)	-0.7463*** (0.1917)	-1.1031*** (0.0648)	-1.1006*** (0.0643)	-1.253*** (0.0715)
<i>ln income</i>	0.7989*** (0.1137)	0.4543*** (0.0458)	0.4979*** (0.0413)	0.4452*** (0.0656)	0.4925*** (0.042)	0.494*** (0.0448)	0.7281*** (0.0906)
<i>food_share</i>	4.6399*** (0.7217)	2.8254*** (0.2056)	3.4482*** (0.2663)	2.5678*** (0.3345)	3.4083*** (0.2704)	3.4142*** (0.2705)	4.6887*** (0.6281)
<i>units</i>	0.2762*** (0.0859)	0.328*** (0.0363)	0.297*** (0.0288)	0.2138*** (0.0748)	0.2766*** (0.0342)	0.2686*** (0.066)	0.332*** (0.0622)
<i>age</i>	-0.0004 (0.003)	0.0013 (0.0027)	0.0006 (0.0011)	0.0004 (0.0027)	0.0005 (0.0012)	0.0002 (0.0013)	0.0012 (0.0023)
<i>children_9</i>	0.0942 (0.0598)	0.0116 (0.0338)	0.0279 (0.0264)	0.1349* (0.0697)	0.0269 (0.0265)	0.0238 (0.0277)	0.0649 (0.0524)
<i>village</i>	0.0467 (0.0716)	0.1552*** (0.0339)	0.1611*** (0.03)	0.2294*** (0.0684)	0.1718*** (0.0319)	0.1621*** (0.036)	0.0719 (0.0629)
<i>woman</i>	0.2573 (0.1766)						
<i>children</i>	0.0789 (0.1067)						
<i>primary</i>	0.1157 (0.1676)						
<i>tertiary</i>	-0.3363*** (0.0928)						
<i>natu_milk</i>	-0.3577** (0.1439)						
σ_v^2	1.8204*** (0.1469)	0.5794*** (0.0311)	0.5833*** (0.0322)	0.6307*** (0.0443)			2.3212*** (0.0701)
ρ				-0.8549*** (0.1022)			
σ_{uv}					-0.4961 (0.4869)		
N	2887	2887	2786	2887	2786	2786	2887
logL	-5912.8	-3630.7	-3202.1	-3643.6	-3201.3	-3199.7 [#]	-5312
n°coef	14	19	9	19	10	14	9
AIC	11853.6	7299.4	6422.2 [#]	7325.2	6422.6	6427.4	10642.0
SBIC	11874.0	7327.1	6435.2 [#]	7352.9	6437.0	6447.6	10655.1

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

N°coef stands for the total number of coefficients estimated by MLE

[#] indicates the model preferred by the given criterion

Table A.4: Poultry

	TOBIT	HURDLE	TWO PART	FIML	LIML	COSS	OLS
<i>constant</i>	-0.8623 (0.5549)	2.9485*** (0.2668)	2.6107*** (0.2688)	1.2198*** (0.2646)	2.7977*** (0.2805)	2.6928*** (0.2826)	0.5027 (0.423)
<i>ln p</i>	-0.9049*** (0.0712)	-0.487*** (0.0722)	-0.6119*** (0.0531)	-0.0513 (0.0756)	-0.6091*** (0.0534)	-0.603*** (0.0538)	-0.885*** (0.0616)
<i>ln income</i>	1.4643*** (0.1441)	0.4912*** (0.0753)	0.7473*** (0.0403)	0.4224*** (0.0759)	0.7455*** (0.0408)	0.7469*** (0.0428)	1.2636*** (0.1055)
<i>food_share</i>	10.4413*** (0.725)	3.7702*** (0.3124)	5.6194*** (0.2556)	1.5139*** (0.3135)	5.5952*** (0.2552)	5.5834*** (0.2562)	9.7036*** (0.5787)
<i>units</i>	0.0941 (0.1172)	0.2583*** (0.0559)	0.1017*** (0.0262)	0.1788*** (0.0567)	0.0601* (0.0318)	-0.0157 (0.0468)	0.2161*** (0.0677)
<i>age</i>	-0.0025 (0.0038)	0.0002 (0.0014)	-0.0018 (0.0011)	0.0052*** (0.0017)	-0.0027** (0.0012)	-0.0031** (0.0013)	0.0007 (0.0028)
<i>children_9</i>	-0.1264* (0.0678)	-0.1396*** (0.0321)	-0.0981*** (0.0263)	-0.0544 (0.061)	-0.1057*** (0.0269)	-0.1142*** (0.0276)	-0.1161** (0.0584)
<i>village</i>	-0.0654 (0.077)	-0.0546 (0.0529)	-0.0032 (0.0281)	0.0564 (0.0692)	0.0054 (0.0289)	0.0304 (0.0314)	-0.0574 (0.0637)
<i>freezer</i>	0.2073 (0.1574)	-0.1024 (0.106)	-0.0026 (0.0525)	0.2746* (0.1452)	-0.0034 (0.0525)	0.0025 (0.0529)	0.196 (0.1347)
<i>woman</i>	1.0786*** (0.2355)						
<i>children</i>	0.103 (0.1346)						
<i>primary</i>	0.7938*** (0.1534)						
<i>tertiary</i>	-0.3745*** (0.1001)						
<i>unemployed</i>	-0.1164 (0.1278)						
<i>natu_poultry</i>	-0.1358 (0.0938)						
σ_v^2	1.9021*** (0.1325)	0.4851*** (0.016)	0.4726*** (0.031)	0.6606** (0.2646)			2.6578*** (0.0527)
ρ				-0.8756*** (0.0756)			
σ_{uv}					-0.4775** (0.2161)		
N	2887	2887	2689	2887	2689	2689	2887
logL	-6440.1	-3532.0	-2808.0	-3648.4	-2804.7	-2796.6 [#]	-5507.5
n°coef	16	21	10	21	11	15	10
AIC	12912.2	7106.0	5636.0	7338.8	5631.4	5623.2 [#]	11035.0
SBIC	12935.6	7136.7	5650.3	7369.5	5647.1	5644.6 [#]	11049.6

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

N°coef stands for the total number of coefficients estimated by MLE

[#] indicates the model preferred by the given criterion

Table A.5: Chocolate

	TOBIT	HURDLE	TWO PART	FIML	LIML	COSS	OLS
<i>constant</i>	-0.9451** (0.4625)	0.5948*** (0.1705)	0.4565 (0.3154)	0.605*** (0.1475)	0.6387* (0.3786)	0.2172 (0.4356)	-0.7416* (0.4038)
<i>ln p</i>	-0.8172*** (0.0649)	-0.1934*** (0.0494)	-0.4922*** (0.0527)	-0.2399*** (0.0501)	-0.4942*** (0.0527)	-0.4948*** (0.0529)	-0.7559*** (0.0601)
<i>ln income</i>	1.37*** (0.1219)	0.3693*** (0.0634)	0.8068*** (0.0579)	0.3786*** (0.0845)	0.7918*** (0.0617)	0.7845*** (0.0631)	1.3261*** (0.1012)
<i>food_share</i>	7.2525*** (0.6643)	1.2819*** (0.3006)	4.6801*** (0.3286)	1.4679*** (0.2587)	4.6636*** (0.3308)	4.6499*** (0.3339)	6.9177*** (0.5818)
<i>units</i>	-0.0897 (0.0928)	0.2107*** (0.0502)	0.0522 (0.04)	0.2716*** (0.0728)	0.0334 (0.043)	0.0564 (0.053)	0.0175 (0.0661)
<i>age</i>	0.0047 (0.0034)	0.0064*** (0.0014)	0.004*** (0.0014)	0.0066*** (0.0016)	0.0034** (0.0015)	0.0038** (0.0016)	0.0072*** (0.0023)
<i>children_9</i>	-0.0049 (0.063)	-0.035 (0.0369)	-0.0124 (0.0333)	-0.0548 (0.0583)	-0.0192 (0.0341)	-0.0127 (0.0349)	0.0365 (0.0553)
<i>village</i>	0.0594 (0.0646)	0.0128 (0.0416)	0.0207 (0.0349)	-0.0026 (0.0434)	0.0201 (0.0349)	0.0206 (0.0349)	0.0506 (0.0579)
<i>woman</i>	0.5505*** (0.1502)						
<i>children</i>	0.1855 (0.1181)						
<i>retired</i>	0.1586* (0.0931)						
<i>primary</i>	-0.4557** (0.2303)						
<i>tertiary</i>	-0.002 (0.0813)						
<i>unemployed</i>	-0.1955 (0.1463)						
σ_v^2	1.3248*** (0.0534)	0.5446*** (0.0225)	0.7335*** (0.0298)	0.7629*** (0.0603)			2.171*** (0.0372)
ρ				-0.4128* (0.2444)			
σ_{uv}					-0.2785 (0.3127)		
N	2887	2887	2618	2887	2618	2618	2887
logL	-6066.9	-3897.7	-3309.1	-4216.2	-3308.7	-3307.4 [#]	-5215.5
n°coef	15	19	9	19	10	14	9
AIC	12163.8	7833.4	6636.2 [#]	8470.4	6637.4	6642.8	10449.0
SBIC	12185.7	7861.1	6649.0 [#]	8498.1	6651.6	6662.7	10462.1

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

N°coef stands for the total number of coefficients estimated by MLE

[#] indicates the model preferred by the given criterion

Table A.6: Pasta

	TOBIT	HURDLE	TWO PART	FIML	LIML	COSS	OLS
<i>constant</i>	-0.7837 (0.6149)	2.7412*** (0.1727)	2.1384*** (0.1963)	3.0078*** (0.5113)	2.1056*** (0.2511)	2.1191*** (0.2082)	0.6752 (0.4125)
<i>ln p</i>	-1.0505*** (0.0756)	-0.3956*** (0.0412)	-0.4771*** (0.0381)	-0.487*** (0.0849)	-0.4779*** (0.0381)	-0.4749*** (0.0381)	-0.8827*** (0.0519)
<i>ln income</i>	1.2079*** (0.1744)	0.1742*** (0.0465)	0.4474*** (0.048)	0.2897*** (0.0693)	0.4498*** (0.0496)	0.4268*** (0.0533)	0.944*** (0.1155)
<i>food_share</i>	10.6715*** (1.0547)	1.9867*** (0.1397)	3.8426*** (0.2725)	2.3068*** (0.3641)	3.8491*** (0.2754)	3.8173*** (0.2768)	8.2085*** (0.7006)
<i>units</i>	0.5348*** (0.1264)	0.336*** (0.0373)	0.187*** (0.0324)	0.1963*** (0.0709)	0.1957*** (0.048)	0.2107*** (0.0749)	0.4458*** (0.0723)
<i>age</i>	-0.0283*** (0.0057)	-0.0083*** (0.0012)	-0.0097*** (0.0011)	-0.0096*** (0.0031)	-0.0098*** (0.0014)	-0.0094*** (0.0014)	-0.0219*** (0.0029)
<i>children_9</i>	-0.2401*** (0.083)	-0.0741** (0.029)	-0.0393 (0.028)	-0.064 (0.0514)	-0.0388 (0.0282)	-0.027 (0.0307)	-0.18*** (0.0582)
<i>village</i>	-0.032 (0.103)	-0.0394 (0.0429)	0.014 (0.0311)	0.1165 (0.0885)	0.0141 (0.0312)	0.0158 (0.0313)	-0.0204 (0.0719)
<i>woman</i>	0.7571*** (0.2726)						
<i>children</i>	-0.084 (0.1683)						
<i>retired</i>	-0.1949 (0.1753)						
<i>primary</i>	0.1215 (0.32)						
<i>tertiary</i>	-0.0956 (0.1294)						
σ_v^2	2.7714*** (0.1543)	0.5031*** (0.0146)	0.5015*** (0.0283)	0.5518*** (0.1426)			3.5737*** (0.0327)
ρ				-0.5125 (0.3884)			
σ_{uv}					0.0714 (0.2999)		
N	2887	2887	2443	2887	2443	2443	2887
logL	-7812.4	-3789.8	-2623.3	-3804.6	-2623.3	-2620.5 [#]	-5934.9
n°coef	14	18	9	18	10	14	9
AIC	15652.8	7615.6	5264.6 [#]	7645.2	5266.6	5269.0	11887.8
SBIC	15673.2	7641.9	5277.1 [#]	7671.5	5280.5	5288.4	11900.9

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

N°coef stands for the total number of coefficients estimated by MLE

[#] indicates the model preferred by the given criterion

Table A.7: Water

	TOBIT	HURDLE	TWO PART	FIML	LIML	COSS	OLS
<i>constant</i>	-5.6209*** (1.3514)	4.6931*** (0.1892)	3.6399*** (0.2427)	4.1472*** (0.5113)	5.1297*** (0.5461)	3.8279*** (0.2782)	-0.4391 (0.6027)
<i>ln p</i>	-2.3659*** (0.1652)	-1.1901*** (0.06)	-1.1865*** (0.0554)	-1.1878*** (0.0849)	-1.1853*** (0.0553)	-1.1835*** (0.055)	-1.7632*** (0.0804)
<i>ln income</i>	2.8788*** (0.35)	0.451*** (0.0498)	0.7184*** (0.0633)	0.5351*** (0.0693)	0.6521*** (0.0672)	0.6927*** (0.0686)	1.6443*** (0.1585)
<i>food_share</i>	13.2427*** (2.2049)	3.0162*** (0.1386)	5.4354*** (0.3858)	3.3221*** (0.3641)	5.3236*** (0.388)	5.3664*** (0.3892)	8.5727*** (1.055)
<i>units</i>	0.283 (0.2999)	0.1908*** (0.0473)	0.0625 (0.0458)	0.1838*** (0.0709)	-0.1173 (0.082)	-0.0329 (0.0713)	0.1415 (0.1108)
<i>age</i>	0.0114 (0.0095)	0.0001 (0.0017)	-0.0004 (0.0016)	0.0027 (0.0031)	-0.0056** (0.0025)	-0.0014 (0.0019)	0.0055 (0.0042)
<i>children_9</i>	-0.3208 (0.2139)	-0.1683*** (0.0439)	-0.1228*** (0.0411)	-0.1247** (0.0514)	-0.1655*** (0.044)	-0.1363*** (0.0427)	-0.2127** (0.1062)
<i>village</i>	0.5441** (0.2336)	0.1806*** (0.0445)	0.2014*** (0.0423)	0.289*** (0.0885)	0.1184** (0.0584)	0.1318** (0.0537)	0.3567*** (0.1167)
<i>woman</i>	-0.2204 (0.5081)						
<i>children</i>	0.0308 (0.3776)						
<i>primary</i>	-0.2438 (0.6961)						
<i>tertiary</i>	-0.6204** (0.287)						
<i>unemployed</i>	-0.5181 (0.4393)						
σ_v^2	11.6547*** (0.6902)	1.0425*** (0.0342)	1.0247*** (0.0302)	1.0314*** (0.1426)			8.0055*** (0.0308)
ρ				0.1167 (0.3884)			
σ_{uv}					-2.0798*** (0.6918)		
N	2887	2887	2409	2887	2409	2409	2887
logL	-9798.3	-4733.9	-3447.6	-4729.4	-3441.1 [#]	-3443.4	-7099.1
n°coef	14	19	9	19	10	14	9
AIC	19624.6	9505.8	6913.2	9496.8	6902.2 [#]	6914.8	14216.2
SBIC	19645.0	9533.5	6925.6	9524.5	6916.0 [#]	6934.1	14229.3

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

N°coef stands for the total number of coefficients estimated by MLE

[#] indicates the model preferred by the given criterion

Table A.8: Sugar

	TOBIT	HURDLE	TWO PART	FIML	LIML	COSS	OLS
<i>constant</i>	-2.5546* (1.3671)	3.9717*** (0.1692)	3.6154*** (0.2126)	3.3483*** (0.8015)	4.5411*** (0.3636)	3.6118*** (0.27)	1.671*** (0.6108)
<i>ln p</i>	-2.644*** (0.2482)	-1.0299*** (0.0486)	-1.0493*** (0.0443)	-0.9527*** (0.1491)	-1.0377*** (0.0446)	-1.041*** (0.0446)	-1.8189*** (0.1195)
<i>ln income</i>	1.4709*** (0.3304)	0.2788*** (0.041)	0.3976*** (0.0488)	0.3003*** (0.1027)	0.4208*** (0.0491)	0.4173*** (0.0489)	0.719*** (0.1464)
<i>food_share</i>	13.914*** (2.1361)	2.3377*** (0.1363)	3.1776*** (0.3426)	2.4514*** (0.4868)	3.1028*** (0.342)	3.1625*** (0.3428)	7.7697*** (0.9981)
<i>units</i>	0.8332*** (0.2228)	0.2299*** (0.0394)	0.1673*** (0.0375)	0.2671*** (0.0658)	0.0422 (0.0541)	0.0919* (0.0508)	0.5166*** (0.1052)
<i>age</i>	0.0146 (0.0109)	0.0061*** (0.0014)	0.0058*** (0.0013)	0.0084* (0.0048)	0.0002 (0.0023)	0.0028 (0.002)	0.012*** (0.0039)
<i>children_9</i>	-0.8608*** (0.2283)	-0.0381 (0.0384)	-0.0169 (0.0369)	-0.051 (0.0815)	-0.0336 (0.0369)	-0.0263 (0.0376)	-0.4321*** (0.1064)
<i>village</i>	0.3972* (0.2271)	0.2767*** (0.0429)	0.2871*** (0.0379)	0.2959*** (0.0914)	0.244*** (0.0473)	0.2552*** (0.0461)	0.3459*** (0.1062)
<i>woman</i>	0.9335** (0.461)						
<i>retired</i>	0.1586 (0.3184)						
<i>primary</i>	0.5173 (0.5645)						
<i>tertiary</i>	-1.0834*** (0.2701)						
<i>agriculture</i>	1.2047* (0.6758)						
σ_v^2	9.5667*** (0.5004)	0.6749*** (0.0188)	0.6732*** (0.0281)	0.7415*** (0.0576)			6.7599*** (0.023)
ρ				0.3459 (0.2667)			
σ_{uv}					-1.2101*** (0.3518)		
N	2887	2887	2210	2887	2210	2210	2887
logL	-9863.6	-4235.7	-2698.5	-4249.6	-2690.0 [#]	-2693.9	-6855.0
n°coef	14	19	9	19	10	14	9
AIC	19755.2	8509.4	5415.0	8537.2	5400.0 [#]	5415.8	13728.0
SBIC	19775.6	8537.1	5427.1	8564.9	5413.4 [#]	5434.6	13741.1

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

N°coef stands for the total number of coefficients estimated by MLE

[#] indicates the model preferred by the given criterion

Table A.9: Potatoes

	TOBIT	HURDLE	TWO PART	FIML	LIML	COSS	OLS
<i>constant</i>	-2.1644 (1.7004)	3.9637*** (0.1918)	3.5463*** (0.2491)	3.1885*** (0.2526)	3.8114*** (0.2563)	3.3027*** (0.3124)	1.8466*** (0.6769)
<i>ln p</i>	-2.7336*** (0.3272)	-1.0076*** (0.0885)	-1.1286*** (0.068)	-0.9608*** (0.0987)	-1.1168*** (0.0674)	-1.1143*** (0.0673)	-1.9993*** (0.1406)
<i>ln income</i>	1.9762*** (0.392)	0.3508*** (0.0574)	0.5457*** (0.0534)	0.4611*** (0.0708)	0.5455*** (0.0533)	0.5447*** (0.0547)	1.1251*** (0.1564)
<i>food_share</i>	19.6421*** (2.3285)	2.8999*** (0.1632)	4.3266*** (0.3569)	2.3697*** (0.1831)	4.1822*** (0.3547)	4.2044*** (0.3587)	11.5125*** (1.0205)
<i>units</i>	-0.4018 (0.2708)	0.2886*** (0.055)	0.1826*** (0.0381)	0.3206*** (0.0878)	0.1623*** (0.0393)	0.1722*** (0.0406)	-0.087 (0.1142)
<i>age</i>	0.0237* (0.0131)	0.0058** (0.0023)	0.0057*** (0.0014)	0.0136*** (0.0021)	0.0047*** (0.0015)	0.005*** (0.0016)	0.0048 (0.0046)
<i>children_9</i>	0.1195 (0.284)	-0.1554*** (0.0593)	-0.1018*** (0.0358)	-0.0335 (0.0657)	-0.0904** (0.0354)	-0.0905** (0.0362)	-0.0819 (0.1223)
<i>village</i>	-2.4084*** (0.2976)	-0.0113 (0.0602)	0.0294 (0.0417)	0.315*** (0.1089)	0.1801*** (0.05)	0.1186* (0.0634)	-1.3477*** (0.1259)
<i>woman</i>	1.25** (0.5278)						
<i>retired</i>	-1.006*** (0.3649)						
<i>primary</i>	1.0531* (0.5478)						
<i>tertiary</i>	-0.1798 (0.3223)						
<i>unemployed</i>	1.3201*** (0.4058)						
<i>agriculture</i>	-2.2992** (1.1609)						
<i>natu_potatoes</i>	-4.0275*** (0.2938)						
σ_v^2	13.6122*** (0.7261)	0.7017*** (0.0268)	0.7051*** (0.0362)	0.8778*** (0.0622)			8.714*** (0.0224)
ρ				-0.6268*** (0.1887)			
σ_{uv}					-0.6016*** (0.1095)		
N	2887	2887	2167	2887	2167	2167	2887
logL	-10001.0	-4148.0	-2696.0	-4174.6	-2677.0	-2676.0 [#]	-7222.0
n°coef	16	21	9	21	10	14	9
AIC	20034.0	8338.0	5410.0	8391.2	5374.0 [#]	5380.0	14462.0
SBIC	20057.4	8368.7	5422.0	8421.9	5387.4 [#]	5398.7	14475.1

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

N°coef stands for the total number of coefficients estimated by MLE

[#] indicates the model preferred by the given criterion

Table A.10: Rice

	TOBIT	HURDLE	TWO PART	FIML	LIML	COSS	OLS
<i>constant</i>	1.1585 (1.2917)	2.8952*** (0.2036)	2.9377*** (0.2392)	2.9703*** (0.4238)	2.4407*** (0.3534)	2.9421*** (0.2551)	2.6959*** (0.5917)
<i>ln p</i>	-2.5805*** (0.2158)	-0.5495*** (0.0502)	-0.7601*** (0.0425)	-0.7821*** (0.1095)	-0.7639*** (0.0426)	-0.7585*** (0.0424)	-1.6325*** (0.0972)
<i>ln income</i>	1.5015*** (0.2778)	0.2554*** (0.0526)	0.4488*** (0.0565)	0.2744*** (0.0556)	0.4589*** (0.0569)	0.4662*** (0.0604)	0.8448*** (0.1273)
<i>food_share</i>	15.5556*** (1.7883)	2.2246*** (0.1383)	3.8016*** (0.3339)	2.1102*** (0.3188)	3.8406*** (0.3332)	3.83*** (0.3364)	8.5285*** (0.8522)
<i>units</i>	0.6627*** (0.2395)	0.1638*** (0.0412)	0.0794** (0.0353)	0.304*** (0.0619)	0.1726*** (0.0635)	0.1059 (0.0711)	0.3976*** (0.0897)
<i>age</i>	-0.0349*** (0.0075)	-0.0044*** (0.0013)	-0.0054*** (0.0012)	-0.0026 (0.0023)	-0.0063*** (0.0015)	-0.0064*** (0.0015)	-0.0172*** (0.0032)
<i>children_9</i>	-0.4793*** (0.1782)	-0.0559 (0.0355)	-0.0375 (0.0333)	-0.0496 (0.048)	-0.026 (0.0342)	-0.0485 (0.0359)	-0.2353*** (0.0888)
<i>village</i>	-0.218 (0.1819)	0.0035 (0.0417)	-0.0064 (0.0336)	0.0092 (0.0824)	-0.0188 (0.0357)	-0.0038 (0.0351)	-0.0954 (0.0871)
<i>woman</i>	0.6504* (0.3865)						
<i>children</i>	0.0253 (0.3116)						
<i>primary</i>	-0.1209 (0.513)						
<i>tertiary</i>	0.0255 (0.2364)						
<i>agriculture</i>	0.8901 (0.5832)						
σ_v^2	6.1621*** (0.3004)	0.4521*** (0.0168)	0.489*** (0.0317)	0.7453*** (0.0885)			5.1725*** (0.0191)
ρ				0.788*** (0.196)			
σ_{uv}					0.6856* (0.3662)		
N	2887	2887	2004	2887	2004	2004	2887
logL	-9472.5	-5978.8	-2126.8	-3883.0	-2125.1	-2124.0 [#]	-6468.7
n°coef	14	19	9	19	10	14	9
AIC	18973.0	11995.6	4271.6	7804.0	4270.2 [#]	4276.0	12955.4
SBIC	18993.4	12023.3	4283.3	7831.7	4283.2 [#]	4294.2	12968.5

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

N°coef stands for the total number of coefficients estimated by MLE

[#] indicates the model preferred by the given criterion

Table A.11: Juices

	TOBIT	HURDLE	TWO PART	FIML	LIML	COSS	OLS
<i>constant</i>	-4.9221* (2.9133)	3.0105*** (0.3115)	2.8649*** (0.3038)	1.625*** (0.6055)	3.3854*** (0.3828)	2.602*** (0.5109)	1.3989* (0.8443)
<i>ln p</i>	-5.1051*** (0.6892)	-0.668*** (0.0906)	-0.7689*** (0.0695)	-0.4783*** (0.1348)	-0.7791*** (0.0697)	-0.7702*** (0.0697)	-2.0642*** (0.2102)
<i>ln income</i>	5.0319*** (0.566)	0.6896*** (0.0595)	0.8089*** (0.0662)	0.7268*** (0.0583)	0.7704*** (0.0682)	0.7874*** (0.0689)	1.8774*** (0.1581)
<i>food_share</i>	29.8197*** (3.4245)	2.3258*** (0.2612)	4.3646*** (0.4257)	1.6294*** (0.5569)	4.3933*** (0.4295)	4.3608*** (0.4314)	9.8824*** (1.004)
<i>units</i>	-0.0869 (0.3967)	-0.0202 (0.0499)	-0.1** (0.0474)	0.1398* (0.0799)	-0.1867*** (0.0637)	-0.1088 (0.0676)	-0.0746 (0.1207)
<i>age</i>	-0.1035*** (0.0185)	-0.0107*** (0.0018)	-0.0134*** (0.0017)	-0.0104*** (0.0023)	-0.0089*** (0.0028)	-0.0122*** (0.0025)	-0.0372*** (0.0041)
<i>children_9</i>	0.0483 (0.3557)	-0.0964** (0.0416)	-0.064 (0.0412)	-0.1036* (0.0615)	-0.0765* (0.041)	-0.0583 (0.0455)	0.0485 (0.1149)
<i>village</i>	-1.6115*** (0.3867)	0.0442 (0.041)	0.0337 (0.0442)	0.1352* (0.0773)	0.0967* (0.053)	0.0513 (0.0522)	-0.4774*** (0.1172)
<i>woman</i>	0.7673 (0.7487)						
<i>retired</i>	-0.681 (0.5753)						
<i>primary</i>	-2.0113** (0.9956)						
<i>tertiary</i>	0.8276* (0.4611)						
<i>unemployed</i>	-0.8214 (0.7592)						
σ_v^2	21.3958*** (0.951)	0.6845*** (0.0225)	0.6959*** (0.0301)	0.9392*** (0.117)			8.0039*** (0.0144)
ρ				0.6217** (0.2675)			
σ_{uv}					-0.6469** (0.3004)		
N	2887	2887	1633	2887	1633	1633	2887
logL	-10221.0	-3902.4	-2021.0	-3943.1	-2018.0 [#]	-2019.0	-7099.0
n°coef	14	19	9	19	10	14	9
AIC	20470.0	7842.8	4060.0	7924.2	4056.0 [#]	4066.0	14216.0
SBIC	20490.4	7870.5	4070.9	7951.9	4068.1 [#]	4083.0	14229.1

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

N°coef stands for the total number of coefficients estimated by MLE

[#] indicates the model preferred by the given criterion

Table A.12: Beef

	TOBIT	HURDLE	TWO PART	FIML	LIML	COSS	OLS
<i>constant</i>	-16.3398*** (3.7367)	0.0041 (0.0631)	-1.5055*** (0.4236)	0.6175*** (0.1712)	-1.8054*** (0.5578)	-1.0414** (0.5115)	-4.4112*** (1.1586)
<i>ln p</i>	-3.5663*** (0.6396)	0.1081*** (0.0403)	0.108 (0.0767)	-0.0159 (0.3593)	0.1065 (0.0767)	0.1061 (0.077)	-1.1592*** (0.2068)
<i>ln income</i>	7.3006*** (0.4779)	0.2869*** (0.0547)	0.6291*** (0.0642)	0.3848 (0.2392)	0.6417*** (0.0667)	0.6424*** (0.0673)	2.4251*** (0.1353)
<i>food_share</i>	35.2727*** (2.7546)	0.7809*** (0.1282)	3.6675*** (0.3756)	1.6691*** (0.2675)	3.6911*** (0.3749)	3.7323*** (0.3801)	12.6048*** (0.8846)
<i>units</i>	-1.467*** (0.4267)	0.1078** (0.0468)	0.0158 (0.0435)	0.0408 (0.3731)	0.0394 (0.0514)	0.0025 (0.0523)	-0.4123*** (0.1066)
<i>age</i>	0.0022 (0.0148)	0.0056*** (0.0017)	0.0057*** (0.0016)	0.0034 (0.2498)	0.0071*** (0.0024)	0.0054** (0.0023)	0.0128*** (0.0033)
<i>children_9</i>	0.0276 (0.3107)	-0.0393 (0.0426)	-0.0315 (0.0404)	-0.0364 (0.4026)	-0.0252 (0.0408)	-0.0281 (0.0406)	-0.0282 (0.0959)
<i>village</i>	-0.6965** (0.3145)	0.0352 (0.0436)	0.0506 (0.0411)	0.0828 (0.3489)	0.0371 (0.0441)	0.0524 (0.0458)	-0.1945* (0.1016)
<i>freezer</i>	0.185 (0.5601)	0.1314 (0.1015)	0.1365* (0.0806)	0.1257 (0.3612)	0.1401* (0.0807)	0.1357* (0.0803)	0.1315 (0.1676)
<i>woman</i>	2.0594*** (0.686)						
<i>children</i>	-0.2227 (0.5327)						
<i>retired</i>	1.3283*** (0.4224)						
<i>primary</i>	-0.9051 (0.9126)						
<i>tertiary</i>	-0.4624 (0.3746)						
<i>unemployed</i>	0.3404 (0.5981)						
<i>natu_beef</i>	-0.6294 (0.4448)						
σ_v^2	14.1764*** (0.6332)	0.5928*** (0.0223)	0.591*** (0.0376)	0.6771 (0.6968)			6.2743*** (0.0141)
ρ				-0.4249 (0.4751)			
σ_{uv}					0.1971 (0.2377)		
N	2887	2887	1535	2887	1535	1535	2887
logL	-9633.7	-3711.0	-1774.5	-3725.0	-1774.0	-1768.9 [#]	-6747.4
n°coef	17	22	10	22	11	15	10
AIC	19301.4	7466.0	3569.0	7494.0	3570.0	3567.8 [#]	13514.8
SBIC	19326.2	7498.1	3580.9 [#]	7526.1	3583.0	3585.6	13529.4

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

N°coef stands for the total number of coefficients estimated by MLE

[#] indicates the model preferred by the given criterion

Table A.13: Fish

	TOBIT	HURDLE	TWO PART	FIML	LIML	COSS	OLS
<i>constant</i>	2.1589 (2.6907)	1.6464*** (0.3716)	3.3992*** (0.3497)	1.0368** (0.4276)	3.0924*** (0.6423)	3.39*** (0.3915)	3.75*** (0.9519)
<i>ln p</i>	-4.21*** (0.4085)	-0.4234*** (0.069)	-0.746*** (0.0523)	-0.4262*** (0.1265)	-0.7448*** (0.0528)	-0.7423*** (0.0526)	-1.9166*** (0.1557)
<i>ln income</i>	3.0483*** (0.419)	0.4144*** (0.0619)	0.5902*** (0.062)	0.4964*** (0.0758)	0.6044*** (0.0643)	0.6217*** (0.0676)	1.2682*** (0.129)
<i>food_share</i>	22.1041*** (2.5154)	1.1998*** (0.2856)	3.0529*** (0.4326)	1.1551*** (0.2886)	3.0664*** (0.4306)	3.0885*** (0.4383)	8.0229*** (0.838)
<i>units</i>	0.4074 (0.3898)	0.0622 (0.0504)	-0.0364 (0.0436)	0.1236 (0.1024)	-0.0076 (0.0683)	-0.009 (0.0697)	0.0658 (0.101)
<i>age</i>	-0.0307** (0.0143)	0.0027 (0.0019)	-0.0047*** (0.0016)	0.0017 (0.0448)	-0.0043** (0.0018)	-0.0043** (0.0018)	-0.0075** (0.0034)
<i>children_9</i>	-0.4268 (0.273)	0.0186 (0.0536)	-0.037 (0.0407)	0.0052 (0.0914)	-0.0298 (0.0429)	-0.0284 (0.0423)	-0.1162 (0.0931)
<i>village</i>	-0.0996 (0.2938)	0.069 (0.0631)	0.0463 (0.0439)	0.0627 (0.0996)	0.042 (0.0447)	0.0387 (0.0455)	-0.0123 (0.0968)
<i>freezer</i>	1.2592** (0.5971)	0.6368*** (0.161)	0.167* (0.0938)	0.4027*** (0.1475)	0.1689* (0.0939)	0.1778* (0.0948)	0.4575** (0.1798)
<i>woman</i>	0.8302 (0.5683)						
<i>children</i>	-0.2863 (0.4979)						
<i>retired</i>	0.239 (0.4212)						
<i>primary</i>	-0.239 (0.7734)						
<i>tertiary</i>	0.6022* (0.336)						
<i>unemployed</i>	-1.108** (0.5264)						
<i>agriculture</i>	0.9052 (0.9251)						
σ_v^2	11.8887*** (0.498)	0.6629*** (0.0262)	0.6473*** (0.0327)	0.9693** (0.4113)			5.8373*** (0.0128)
ρ				0.6807** (0.3085)			
σ_{uv}					0.2352 (0.3931)		
N	2887	2887	1516	2887	1516	1516	2887
logL	-9678.7	-3787.6	-1821.4	-3827.2	-1821.2	-1818.5 [#]	-6643.2
n°coef	17	22	10	22	11	15	10
AIC	19391.4	7619.2	3662.8 [#]	7698.4	3664.4	3667.0	13306.4
SBIC	19416.2	7651.3	3674.6 [#]	7730.5	3677.4	3684.7	13321.0

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

N°coef stands for the total number of coefficients estimated by MLE

[#] indicates the model preferred by the given criterion

Table A.14: Legumes

	TOBIT	HURDLE	TWO PART	FIML	LIML	COSS	OLS
<i>constant</i>	0.3153 (4.5408)	1.0474*** (0.2045)	1.8331*** (0.3521)	0.4601*** (0.0627)	1.8051** (0.8039)	1.7128*** (0.4041)	3.0635*** (1.1788)
<i>ln p</i>	-6.5334*** (1.0369)	-0.168*** (0.0555)	-0.4047*** (0.0585)	-0.3747 (0.4111)	-0.4045*** (0.0589)	-0.416*** (0.0598)	-2.0631*** (0.2832)
<i>ln income</i>	2.7136*** (0.5042)	0.1973*** (0.0551)	0.2413*** (0.0766)	0.0859 (0.2574)	0.2416*** (0.0767)	0.2342*** (0.0794)	0.6284*** (0.1052)
<i>food_share</i>	19.2672*** (3.2822)	0.8309*** (0.1582)	2.3568*** (0.4329)	0.4591*** (0.0819)	2.3578*** (0.4349)	2.2992*** (0.4412)	4.5218*** (0.7368)
<i>units</i>	0.8984** (0.3881)	0.0714* (0.042)	0.0281 (0.0467)	0.3118 (0.2378)	0.0311 (0.0923)	0.016 (0.0862)	0.2573*** (0.0926)
<i>age</i>	-0.012 (0.0171)	0.0035* (0.0018)	-0.0007 (0.0017)	0.0061 (0.2363)	-0.0006 (0.0023)	-0.001 (0.002)	0.0002 (0.0026)
<i>children_9</i>	-0.3953 (0.3197)	-0.0459 (0.0383)	-0.0435 (0.0403)	0.0485 (0.3025)	-0.0428 (0.0443)	-0.0382 (0.0438)	-0.0843 (0.0769)
<i>village</i>	-0.4778 (0.3769)	0.0841* (0.0492)	0.0772* (0.0454)	0.0469 (0.282)	0.0764 (0.0486)	0.0855* (0.0499)	-0.0899 (0.084)
<i>woman</i>	1.8245*** (0.7001)						
<i>retired</i>	0.5233 (0.4949)						
<i>primary</i>	-0.6339 (1.0209)						
<i>tertiary</i>	0.6211 (0.4417)						
<i>agriculture</i>	0.7725 (1.1041)						
σ_v^2	16.2787*** (0.6842)	0.3587*** (0.0189)	0.3649*** (0.0495)	1.155*** (0.2001)			4.0063*** (0.0226)
ρ				0.9497*** (0.3188)			
σ_{uv}					0.0152 (0.3799)		
N	2887	2887	844	2887	844	844	2887
logL	-7429.5	-2458.4	-772.2	-2465.0	-772.2	-767.2 [#]	-6099.9
n°coef	14	19	9	19	10	14	9
AIC	14887.0	4954.8	1562.4 [#]	4968.0	1564.4	1562.4 [#]	12217.8
SBIC	14907.4	4982.5	1570.7 [#]	4995.7	1573.7	1575.4	12230.9

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

N°coef stands for the total number of coefficients estimated by MLE

[#] indicates the model preferred by the given criterion

Table A.15: Cocoa

	TOBIT	HURDLE	TWO PART	FIML	LIML	COSS	OLS
<i>constant</i>	8.2355*** (2.6349)	0.9561*** (0.3491)	5.4474*** (0.4494)	1.1133*** (0.0721)	4.2036*** (1.1177)	5.3766*** (0.4829)	8.2747*** (1.2554)
<i>ln p</i>	-3.7465*** (0.4822)	-0.2934*** (0.0773)	-1.2194*** (0.0723)	-0.8225*** (0.1815)	-1.2198*** (0.0721)	-1.2146*** (0.0723)	-2.4392*** (0.2366)
<i>ln income</i>	1.005*** (0.26)	0.3037*** (0.0818)	0.3595*** (0.0906)	0.4171*** (0.0513)	0.3489*** (0.0901)	0.3967*** (0.0948)	0.3307*** (0.0885)
<i>food_share</i>	7.053*** (1.5328)	0.4375* (0.2295)	2.0757*** (0.5609)	0.7311*** (0.0382)	2.1416*** (0.5643)	2.2443*** (0.5689)	2.7959*** (0.5807)
<i>units</i>	0.7697*** (0.2209)	0.1891*** (0.0655)	0.0785 (0.0607)	0.415*** (0.0764)	0.2542 (0.1571)	0.0574 (0.1258)	0.33*** (0.0721)
<i>age</i>	0.0054 (0.007)	0.0013 (0.0026)	-0.006*** (0.0022)	0.0033 (0.0932)	-0.0042 (0.003)	-0.0048* (0.0028)	0.0016 (0.0022)
<i>children_9</i>	0.2043 (0.1594)	0.0109 (0.0559)	0.0746 (0.049)	0.0869 (0.0855)	0.1011* (0.0526)	0.0747 (0.0536)	0.0758 (0.0685)
<i>village</i>	0.1747 (0.1725)	-0.0246 (0.0562)	0.0466 (0.0463)	0.1692** (0.0843)	0.0757 (0.0527)	0.0419 (0.0524)	0.0819 (0.0672)
<i>woman</i>	-0.0289 (0.3319)						
<i>children</i>	-0.0599 (0.2751)						
<i>primary</i>	0.2363 (0.4631)						
<i>tertiary</i>	-0.5063** (0.2015)						
σ_v^2	3.5766*** (0.2435)	0.3446*** (0.0266)	0.4898*** (0.051)	1.2419*** (0.0383)			2.6126*** (0.0237)
ρ				0.8617*** (0.0331)			
σ_{uv}					0.6824 (0.5206)		
N	2887	2887	809	2887	809	809	2887
logL	-6623.8	-1986.6	-859.2	-2553.0	-858.1	-851.3 [#]	-5482.7
n°coef	13	18	9	18	10	14	9
AIC	13273.6	4009.2	1736.4	5142.0	1736.2	1730.6 [#]	10983.4
SBIC	13292.6	4035.5	1744.6	5168.3	1745.3	1743.3 [#]	10996.5

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

N°coef stands for the total number of coefficients estimated by MLE

[#] indicates the model preferred by the given criterion

Acknowledgement

This research was supported by the Charles University, project GA UK No. 286517. Further, this research has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 681228 (GEMCLIME). This support is gratefully acknowledged. Responsibility for any errors remains with the authors.

IES Working Paper Series

2017

1. Petra Lunackova, Jan Prusa, Karel Janda: *The Merit Order Effect of Czech Photovoltaic Plants*
2. Tomas Havranek, Zuzana Irsova, Tomas Vlach: *Measuring the Income Elasticity of Water Demand: The Importance of Publication and Endogeneity Biases*
3. Diana Zigraiova, Petr Jakubik: *Updating the Long Term Rate in Time: A Possible Approach*
4. Vaclav Korbel, Michal Paulus: *Do Teaching Practices Impact Socio-emotional Skills?*
5. Karel Janda, Jan Malek, Lukas Recka: *Influence of Renewable Energy Sources on Electricity Transmission Networks in Central Europe*
6. Karel Janda, Jan Malek, Lukas Recka: *The Influence of Renewable Energy Sources on the Czech Electricity Transmission System*
7. Ondrej Filip, Karel Janda, Ladislav Kristoufek: *Prices of Biofuels and Related Commodities: An Example of Combined Economics and Graph Theory Approach*
8. Adam Kucera: *Interest Rates Modeling and Forecasting: Do Macroeconomic Factors Matter?*
9. Arshad Hayat: *Foreign Direct Investments, Institutional Framework and Economic Growth*
10. Jana Votapkova, Pavlina Zilova: *Health Status as a Determinant for Pre-Retirement Savings*
11. Matej Nevrla: *Systemic Risk in the European Financial and Energy Sector: Dynamic Factor Copula Approach*
12. Mojmir Hampl, Tomas Havranek: *Should Inflation Measures Used by Central Banks Incorporate House Prices? The Czech Approach*
13. Radek Janhuba, Kristyna Cechova: *Criminals on the Field: A Study of College Football*
14. Anton Astakhov, Tomas Havranek, Jiri Novak: *Firm Size and Stock Returns: A Meta-Analysis*
15. Martin Stepanek: *Pension Reforms and Adverse Demographics: The Case of the Czech Republic*
16. Tomas Havranek, Zuzana Irsova, Olesia Zeynalova: *Tuition Fees and University Enrollment: A Meta-Analysis*
17. Oliver Polyak, Jitka Postulkova: *Central Eastern and South Eastern European Markets Macro-Fundamental Analysis*
18. Jiri Witzany: *A Bayesian Approach to Backtest Overfitting*
19. Jozef Barunik, Lucie Kraicova: *Common Cycles in Volatility and Cross Section of Stock Returns*
20. Frantisek Cech, Jozef Barunik: *Measurement of Common Risk Factors: A Panel Quantile Regression Model for Returns*

21. Sarlota Smutna, Milan Scasny: *Selectivity Problem in Demand Analysis: Single Equation Approach*

All papers can be downloaded at: <http://ies.fsv.cuni.cz>



Univerzita Karlova v Praze, Fakulta sociálních věd

Institut ekonomických studií [UK FSV – IES] Praha 1, Opletalova 26

E-mail : ies@fsv.cuni.cz

<http://ies.fsv.cuni.cz>