Censored QUAIDS estimation with *quaidsce*[*]

Juan Carlos Caro, University of Luxembourg, Luxembourg

Grace Melo, Texas A&M University, US

José Alberto Molina, University of Zaragoza, Spain, and IZA, Germany

Juan Carlos Salgado, Instituto Nacional de Salúd Pública, México

December 12, 2021

Abstract

Censoring, or zero expenditures, in the dependent variables of demand systems can lead to inconsistent parameter and elasticity estimates. Poi (2012) introduced the Stata command *quaids* to estimate quadratic almost-ideal demand systems, or QUAIDS (Banks, Bundell and Lewbel, 1997), although without the possibility to address censoring. In this paper, we introduce the command *quaidsce* to consistently estimate the QUAIDS in the presence of zero observations, using the two-step procedure proposed by Shonkwiler and Yen (1999). The new command also allows for estimating expenditure and price elasticities with standard errors, using post-estimation tools.

Keywords: Censoring, Stata, QUAIDS, *quaidsce*

JEL codes: D12, C30, C87

---

[*] Correspondence to Juan Carlos Caro. Department of Cognitive and Behavioral Sciences, University of Luxembourg. Email: juan.caro@uni.lu

Acknowledgment: This paper was partially written while Jose Alberto Molina was Visiting Fellow at the Department of Economics of the Boston College (US), to which he would like to express his thanks for the hospitality and facilities provided. It has benefitted from funding from the Government of Aragón (Project S32-20R).

1. Introduction

Censoring, or the presence of zero expenditures, in the dependent variables of demand systems has been an important topic in economics and econometrics for decades (Houthakker, 1953; Deaton, 1986). In demand analyses, a high proportion of zero expenditures in the budget equations makes it challenging to estimate theoretically consistent consumer demand models at a disaggregated level. Among the different theoretical approaches in the economic literature, Shonkwiler and Yen (1999) propose a consistent two-step procedure to a demand system model, introduced by Banks et al. (1997). From a practical standpoint, the user-written commands for AIDS estimation, namely **quaids** by Poi (2012) and **aidsills** by Lecocq and Robin (2015), do not account for censoring. Despite the high level of zero observations in micro-data analyses, particularly food demand systems, there is no command in Stata available to account for censoring in AIDS estimation. We bridge this gap by introducing the command **quaidsce** to consistently estimate the two-stage AIDS in the presence of zero observations. The new command also estimates the expenditure and price elasticities.

2. Censored QUAIDS model

The standard QUAIDS can be specified as follows:

$$w_{hi} = \alpha_i + \sum_{j=1}^{J} \gamma_{ij} \ln p_{hj} + \beta_i \{\ln y_h - \ln a(\mathrm{p})\} + \lambda_i \frac{\{ln\, y_h - \ln a(\mathrm{p})\}^2}{b(\mathrm{p})} + \sum_{k=1}^{K} \eta_{ik} z_{hk} + u_{hi}$$

$$h = 1, \dots, H;\; i = 1, \dots, J; k = 1, \dots, K \qquad (2)$$

where w_{hi} is the budget share of category i for household h, p_{hj} is the price paid by household h for goods from category j, y_h is total household expenditure, z_{hk} is the k^{th} demand shifter, and u_{hi} is the residual. The terms *H, J,* and *K* indicate the number of households, categories, and shifters, respectively. The corresponding model parameters are $\alpha_i, \gamma_{ij}, \beta_{i,}\lambda_i$, and η_{ik}. The price aggregators

$a(\text{p})$ and $b(\text{p})$ can be defined by $\ln a(\text{p}) = \alpha_o + \sum_{j=1}^{J} \alpha_{ij} \ln p_{hj} + \sum_{j=1}^{J} \sum_{k=1}^{K} \eta_{ik} z_{hk} \ln p_{hj} + 0.5 \sum_{i=1}^{J} \sum_{j=1}^{J} \eta_{ij} \ln p_{hi} \ln p_{hj}$ and $b(\text{p}) = \exp(\sum_{j=1}^{J} \beta_i \ln p_{hj})$.

Consistent with theory, three sets of restrictions are imposed: $\gamma_{ij} = \gamma_{ji}$ for all $i \neq j$ (symmetry); $\sum_{i=1}^{J} \alpha_i = 1$, $\sum_{i=1}^{J} \lambda_i = 0$, $\sum_{i=1}^{J} \beta_i = 0$, and $\sum_{i=1}^{J} \eta_{ik} = 0, \forall\, k$ (adding up); and $\sum_{j=1}^{J} \gamma_{ij} = 0, \forall\, j$ (homogeneity). The adding up restriction $\sum_{i=1}^{J} w_i = 1$ was not imposed; therefore, the demand equation for the J good is estimated as in a nonlinear system of equations. Under this two-step approach, total budget shares are not required to sum to unity when taking censoring into account (Yen et al., 2003). For the estimation, the approach suggested by Deaton and Muellbauer (1980) was followed and α_o was set equal to a constant. The latent budget share w_{hi}^* is related to the observed budget share, as follows: $w_{hi} = d_{hi} w_{hi}^*$, where w_{hi} is calculated as the category-level expenditure divided by total weekly expenses, and d_{hi} is a binary dependent variable that equals 1 for nonzero expenditures and 0 otherwise. Following Shonkwiler and Yen (1999), the unconditional expected value of the system in (2) can be written as $w_{hi}^* = \Phi(z_h' \theta_i)\, w_{hi} + \delta_i\, \phi(z_h{}' \theta_i) + \varepsilon_i$, where Φ and ϕ are the cumulative and density distribution functions for a standard normal variable, respectively, which can be estimated by a two-step approach using all observations.[1] First, a univariate probit equation $d_{hi} = z_h{}' \theta_i \;\forall\, i$ is estimated for all categories, where z_h is a vector of regressors that includes category prices and a set of demographic variables. Second, $\Phi(z_h' \widehat{\theta_\iota})$ and $\phi\big(z_h' \widehat{\theta_\iota}\big)$ are calculated and included in the second step as follows, with the observed budget shares in (3) modeled as in (2):

$$w_{hi}^* = \Phi(z_h' \widehat{\theta_\iota})\, w_{hi} + \delta_i \phi\big(z_h' \widehat{\theta_\iota}\big) + \varepsilon_i \qquad (3)$$

In addition to censoring, another potential source of endogeneity in equation (2) comes from total expenditures since a change in budget share allocations is likely to affect the total budget. The common solution to this issue is to use disposable household income and its quadratic

[1] To estimate a censored demand system, one can alternatively assume that the errors in the first and second steps follow a joint normal distribution, and estimate the demand system with maximum likelihood estimation (MLE). This approach, however, relies on a strong distributional assumption about the residuals, which may not hold. In addition, estimation with MLE can result in computationally burdensome demand systems.

term as instruments in a two-stage least squares (2SLS) type of estimator for a system of equations (see, e.g., Blundell and Robin, 1999). In the first stage, total expenditure is regressed on the exogenous control variables and the instruments. Then, the residuals from this regression are added to every equation in the system via (2) as additional control variables. Blundell and Robin (1999) show that under the assumption that the error term u from (2) can be orthogonally decomposed into the residuals from stage one and a white noise term, the augmented regression estimator is identical to the 2SLS estimator. Thus, one could estimate the extended OLS estimator to account for total expenditures endogeneity in (2).

Demand elasticity formulas that account for the influence of the probability terms in (3) are presented next. Omitting the *h* subscripts, differentiating (3) with respect to *y* and *p* yields:

$$\mu_i = \beta_i + 2\lambda_i \frac{\{\ln y - \ln a(\mathrm{p})\}}{b(\mathrm{p})} \tag{4}$$

$$\mu_{ij} = \gamma_{ij} - \mu_i \left(\alpha_j + \sum_{j=1}^{J} \gamma_{ij} \ln p_{hj}\right) - \lambda_i \, \beta_j \frac{\{\ln y - \ln a(\mathrm{p})\}^2}{b(p)} \tag{5}$$

Expenditure elasticities are then given by $e_i = 1 + \frac{1}{w_i^*} \left\{\Phi_\mathrm{i} \, \mu_{ij} + \theta_{i,\ln(p_j)} \phi_i [w_i - \delta_i(z'\theta_i)]\right\}$, and uncompensated price elasticities by $e_{ij}^{\mathrm{u}} = -\,\delta_{ij} + \frac{1}{w_i^*} \left\{\Phi_\mathrm{i} \, \mu_{ij} + \theta_{i,\ln(p_j)} \phi_i [w_i - \delta_i(z'\theta_i)]\right\}$, where δ_{ij} is the Kronecker delta, which equals 1 if $i = j$ and 0 otherwise.

It is worth noting that as $\Phi_\mathrm{i} \rightarrow 1$ and $\phi_i \rightarrow 0$,, the elasticity formulas collapse into the standard elasticity formulas for QUAIDS, which does not address censoring.

Estimation is conducted by iterative feasible generalized nonlinear least-squares (ifgnls) using the **nlsur** command, extending from the **quaids** command developed by Poi (2012). The default for the variance-covariance matrix is the asymptotic theory consistent (**gnr**). When censoring is not selected, the **quaidsce** and **quaids** commands produce equivalent results, by excluding the last equation to avoid singularity in the error covariance matrix. The function evaluator, as well as the routines for post-estimation, are included in .ado and .mata files that are installed along with the command.

3. The quaidsce command

The **quaidsce** command syntaxis for a flexible AIDS model, with or without demographics, censoring, and quadratic term, follows:

```
quaidsce varlist_expshares [if] [in], anot(#)
{prices(varlist_prices)|lnprices(varlist_lnprices)}
{expenditure(varlist_exp)|lnexpenditure(varlist_lnexp)}
[demographics(varlist_demo) noquadratic nocensor nolog vce(vcetype)
method(method_name) level(#)]
```

`varlist_expshares` denotes the list of variables containing the expenditure shares for the number of goods in the demand system. For each individual, shares should add up to unity, regardless of the approach used, **`quaidsce`** will adjust the parameter restrictions suited to each model (e.g. censored versus uncensored). A vector of prices (matching the dimension of the number of goods) must be provided, either in levels (`varlist_prices`) or in natural logarithms (`varlist_lnprices`), following the same order as `varlist_expshares`. Total expenditure must also be provided in levels or natural logarithm using `expenditure()` or `lnexpenditure()` option (only one of them). Total expenditure should be $\sum_{i=1}^{J} p_i q_i = y_h$ for each observation.

A value for α_o must be provided using `anot()` Additional demographic variables can be included using the option `demographics()`. Using the options `noquadratic` and `nocensor`, **quaidsce** can accommodate different models. Iteration log can be suppressed using `nolog`. The option `level(#)` sets the level for the confidence intervals. The `vce(vcetype)` and `method(method_name)` options control the type of method for the variance-covariance matrix estimation and the estimator used for computation, using the same options as `nlsur`; see [R] **`nlsur`**. By default, **quaidsce** estimates a demand system via ifgnls using the vce(gnr) option. See the technical notes section for additional comments on estimation approaches.

After estimation, the `predict` command obtains expenditure shares based on the model parameters, providing stub names for the new generated variables (creating *J* variables using the same stub name). By default, predictions will be done for all observations in the dataset. The syntaxis follows:

```
predict {stub*} [if] [in]
```

Similarly, `quaidsce` post-estimation allows for direct computation of the income and price elasticities (and standard errors) at the sample means, using the `estat` commands. The output in each case is two matrices of the same dimension, including the mean estimates and the standard errors respectively, using the delta method. In each case, the `estat` command will use the specification of the estimated model to compute the elasticity using the corresponding formula, as noted above. For each type of elasticity (expenditure, compensated price, and uncompensated price), the syntaxis is described below.

estat **expenditure** [if] [in]

estat **compensated** [if] [in]

estat **uncompensated** [if] [in]

The estimated matrices follow the same order as the order of variables introduced in the model estimation. For example, the element in row *i*, column *j* of the matrices produced after `estat compensated` contains the compensated price elasticity of good *i* with respect to the price of good *j*.

Saved results in `e()` for `quaidsce` are:

```
Scalars
  e(N)                number of observations
  e(ll)               log likelihood
  e(N_clust)          number of clusters
  e(ndemos)           number of demographic variables
  e(anot)             value of alpha_0
  e(ngoods)           number of goods
```

```
Macros
     e(cmd)              quaidsce
     e(clustvar)         name of cluster variable
     e(vce)              vcetype specified in vce()
     e(vcetype)          title used in label Std. Err.
     e(properties)       b V
     e(estat_cmd)        program used to implement estat
     e(predict)          program used to implement predict
     e(demographics)     demographic variables included
     e(lhs)              expenditure share variables
     e(expenditure)      expenditure variable
     e(lnexpenditure)    log-expenditure variable
     e(prices)           price variables
     e(lnprices)         log-price variables
     e(quadratic)        noquadratic
     e(censor)           nocensor
     e(method)           specified in method()
     e(properties)       b V

   Matrices
     e(b)                coefficient vector
     e(V)                variance-covariance matrix of the estimators
     e(best)             coefficient vector of estimated parameters
     e(Vest)             variance-covariance matrix of estimated parameters
     e(alpha)            alpha vector
     e(beta)             beta vector
     e(gamma)            gamma matrix
     e(lambda)           lambda vector
     e(eta)              eta matrix
     e(rho)              rho vector
     e(delta)            delta vector

   Functions
     e(sample)           marks estimation sample
```

The post-estimation commands produce the following stored matrices:

```
r(elas_i)         vector of expenditure elasticities
r(se_elas_i)      vector of s.e. of expenditure elasticities
r(elas_c)         matrix of compensated price elasticities
r(se_elas_c)      matrix of s.e. of compensated price elasticities
r(elas_u)         matrix of uncompensated price elasticities
r(se_elas_u)      matrix of s.e. of uncompensated price elasticities
```

4. Application

We illustrate the use of **quaidsce** and its companion post-estimation commands by estimating a food demand system using expenditures data from a nationally representative survey. We fit a censored QUAIDS model for 17 food-at-home categories with varying censoring rates, using data from the Household Budget Survey (EPF, Spanish acronym), collected by the Chilean National Institute of Statistics for the 2016/2017 period (INE, 2020). The data were collected from a sample of households using self-reported diaries of all purchases, including food, over two weeks. Data include monthly income and expenditure values of 15,147 households. Quantity information was requested from INE to calculate quality-adjusted unit values based on the approach of Crawford et al. (2003) and later adapted by Capacci and Mazzocchi (2011), which were used as proxies of prices.

EPF 2016/2017 Descriptive Statistics

Group	Purchase > 0	Quantity (gr/day/capita)	Expenditure Shares
1 Starches	0.635	89.63	0.033
2 Bread	0.968	197.82	0.148
3 Breakfast cereals	0.264	20.25	0.009
4 Unprocessed meat	0.887	146.69	0.199
5 Processed meat	0.824	40.89	0.068
6 Milk and dairy desserts	0.733	164.23	0.058
7 Cheese	0.707	25.79	0.043
8 Fruits	0.685	245.62	0.045
9 Vegetables	0.891	212.15	0.121
10 Legumes & proc. FVs	0.543	24.07	0.024
11 Sweets	0.587	36.81	0.032
12 Snacks	0.750	38.30	0.062
13 Unsweetened beverages	0.580	882.32	0.031
14 Sweetened beverages	0.854	287.12	0.089
15 Fats & oils	0.635	43.73	0.028
16 Refined sugar	0.328	63.31	0.008
17 Nonsugar sweeteners	0.073	7.93	0.002

On average, there is a wide range of censoring, particularly for groups that also represent a small fraction of food expenditures. As zero expenditures in a category may reflect not only low preferences but also unaffordability, unavailability, or low purchase frequencies, censoring represents an important source of bias in our application of demand analysis. To address this bias, we estimated a censored model using **quaidsce**. For comparison, we also estimated an uncensored model, using the same demographic and control variables, with anot(10):

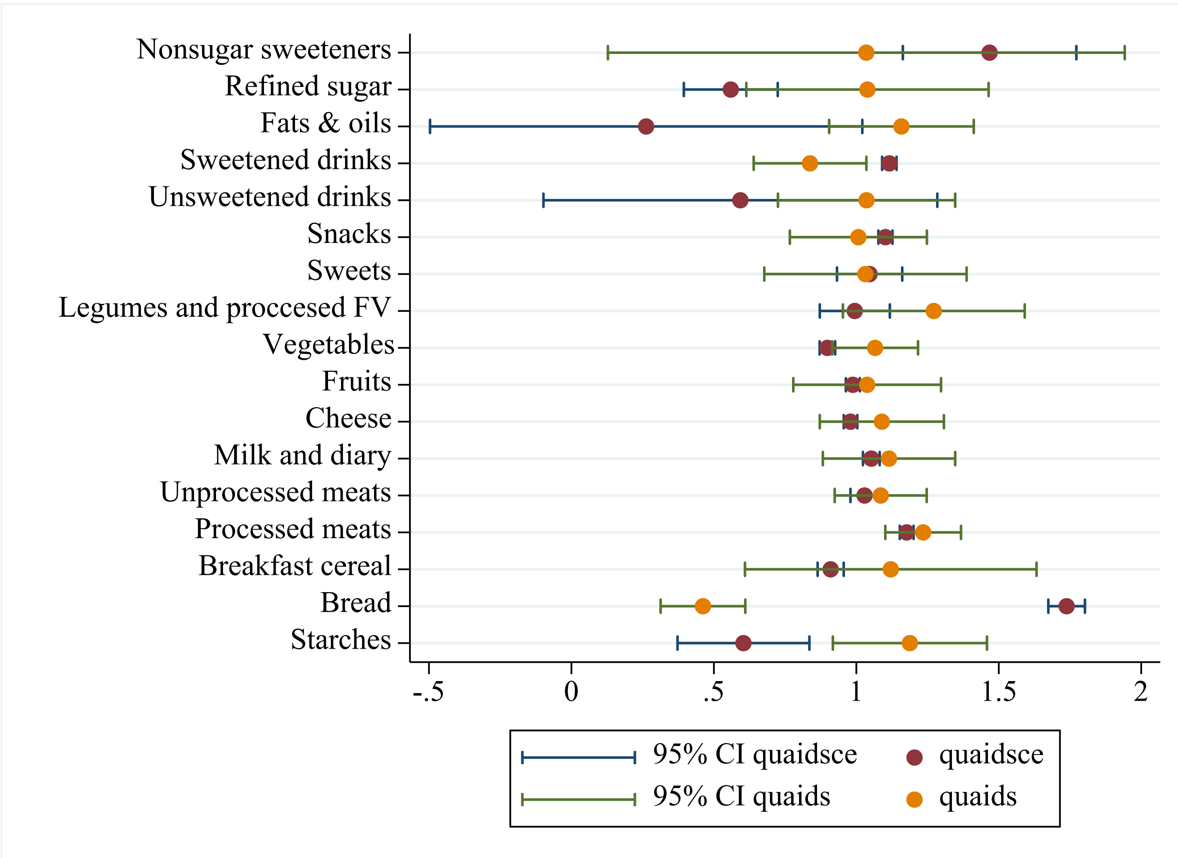
```
quaidsce w1-w17, anot(10) prices(p1-p17) expenditure(total) nolog demographics(x1-x3)
```

```
(obs = 15,147)
Calculating NLS estimates...
Calculating FGNLS estimates...
[output ommited]

Censored Quadratic AIDS model
--------------------
Number of obs          =      15147
Number of demographics =          3
Alpha_0                =         10
Log-likelihood         =  482472.11

------------------------------------------------------------------------------
             | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
alpha        |
     alpha_1 |  -.0123345   .0071769    -1.72   0.086    -.0264009    .0017319
     alpha_2 |  -.5565543   .0088003   -63.24   0.000    -.5738026    -.539306
     alpha_3 |   .0551545   .0036256    15.21   0.000     .0480485    .0622606
     alpha_4 |   .3217646   .0113305    28.40   0.000     .2995573     .343972
     alpha_5 |  -.1923305   .0104614   -18.38   0.000    -.2128344   -.1718265
     alpha_6 |   .1132634   .0091764    12.34   0.000     .0952781    .1312487
     alpha_7 |  -.0214901   .0089209    -2.41   0.016    -.0389746   -.0040055
     alpha_8 |   .1165358   .0068222    17.08   0.000     .1031647     .129907
     alpha_9 |   .1315745   .0121268    10.85   0.000     .1078063    .1553427
    alpha_10 |   .0112371   .0062214     1.81   0.071    -.0009566    .0234308
    alpha_11 |   .0228944   .0054697     4.19   0.000      .012174    .0336149
    alpha_12 |   .4783736   .0090051    53.12   0.000     .4607239    .4960233
    alpha_13 |   .0190407   .0040675     4.68   0.000     .0110686    .0270129
    alpha_14 |   .3570162   .0078524    45.47   0.000     .3416258    .3724066
    alpha_15 |   -.021912   .0066388    -3.30   0.001    -.0349237   -.0089002
    alpha_16 |   .0028298   .0035388     0.80   0.424    -.0041062    .0097659
    alpha_17 |   .0084027   .0022578     3.72   0.000     .0039776    .0128279
-------------+----------------------------------------------------------------
[output ommitted]
-------------+----------------------------------------------------------------
```

```
delta        |
     delta_1 |   .0419601   .0020001    20.98   0.000      .03804    .0458801
     delta_2 |   .3025808   .0021382   141.51   0.000    .2983901    .3067715
     delta_3 |   .0191154   .0009176    20.83   0.000     .017317    .0209138
     delta_4 |   .1530532   .0022877    66.90   0.000    .1485695     .157537
     delta_5 |   .0459996   .0022873    20.11   0.000    .0415167    .0504826
     delta_6 |   .1467501   .0027192    53.97   0.000    .1414206    .1520796
     delta_7 |   .1138754   .0022876    49.78   0.000    .1093919     .118359
     delta_8 |    .066189   .0018492    35.79   0.000    .0625646    .0698134
     delta_9 |   .3052717     .00332    91.95   0.000    .2987646    .3117789
    delta_10 |  -.0736759   .0020962   -35.15   0.000   -.0777844   -.0695675
    delta_11 |   .0407077   .0012353    32.95   0.000    .0382865    .0431289
    delta_12 |   .0739183   .0019064    38.77   0.000    .0701818    .0776548
    delta_13 |   .0331155   .0014532    22.79   0.000    .0302672    .0359637
    delta_14 |   .0642146   .0014937    42.99   0.000     .061287    .0671423
    delta_15 |  -.0183257   .0021159    -8.66   0.000   -.0224729   -.0141786
    delta_16 |  -.0213836   .0010089   -21.19   0.000    -.023361   -.0194061
    delta_17 |  -.0110664   .0009962   -11.11   0.000    -.013019   -.0091139
-------------+----------------------------------------------------------------
[output ommitted]
```

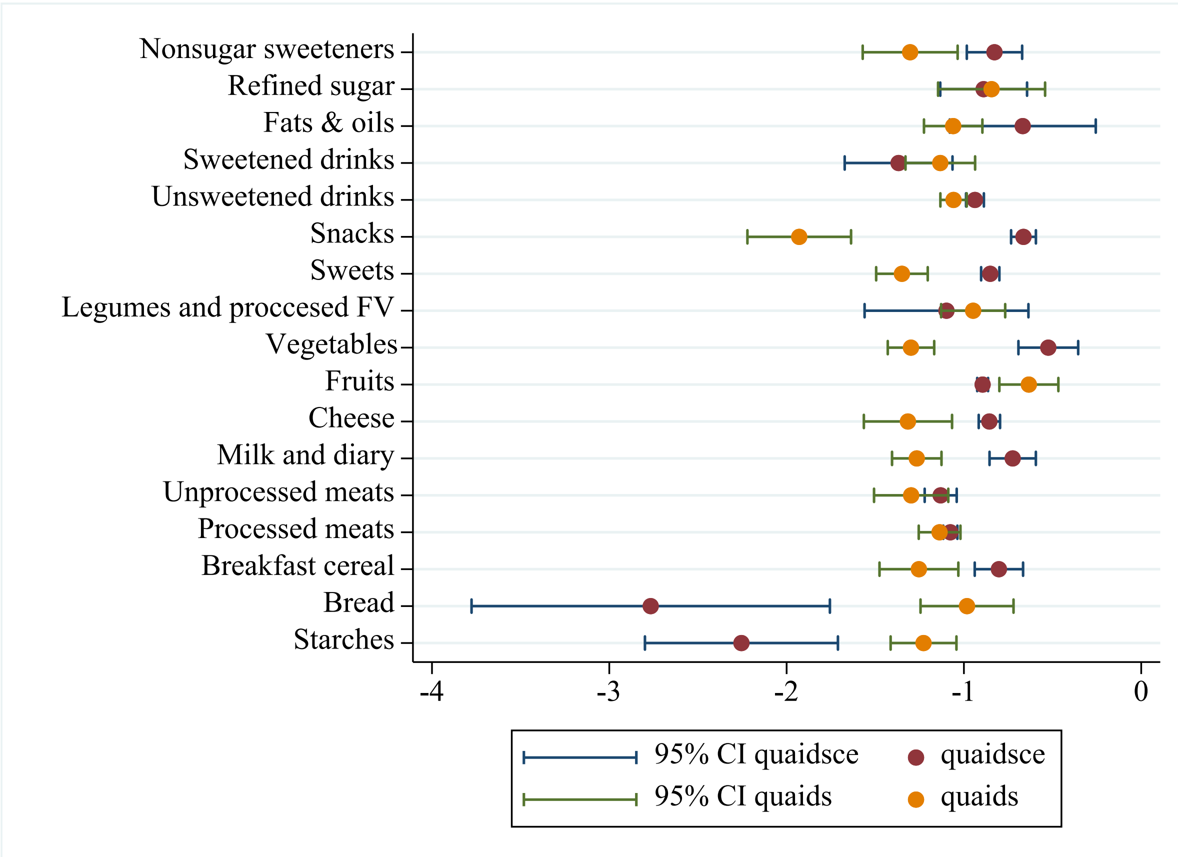
We noted that in every equation, the parameter that accounts for the correction due to zero consumption (delta) is significantly different from zero. Positive values for delta imply an inverse relationship between the likelihood of spending and the average expenditure share for a given food group (the inverse is true for negative values).

Fig 1: Expenditure elasticities



After estimation, we use the `estat` command to produce expenditure and uncompensated price elasticities. Figures 1 and 2 show the differences between the estimated elasticities with and without correction due to censoring (with 95% confidence intervals). In general, there are important differences in the mean estimated elasticities when censoring is taken into account, changing the interpretation of results dramatically. Overall, we show that ignoring the potential bias of zero expenditures can lead to inaccurate estimates of demand responses to prices and income changes, and therefore affecting inferences in policy analysis.

Fig 2: Uncompensated own price elasticities



5. Technical notes

The **quaidsce** command was developed as an extension to the **quaids** command, in order to provide a direct solution to censoring in estimation of demand systems. As such, we do not directly address the potential endogeneity of prices and total expenditures. We recommend researchers use the methods described in this paper to address price and expenditure endogeneity to the extent the data permits. For instance, to address total expenditures endogeneity, researchers could use the control function approach discussed in Banks et al. (1997).

We recommend using bootstrap methods for standard errors in the censored model estimation due to the nonlinear nature of the model. This is particularly important, as it will be used as an input to estimate the standard errors of the predicted elasticities. Similarly, we only recommend making inferences when the models are estimated using the ifgnls method. The `method(method_name)` is added for experienced users interested in debugging when the model cannot be fitted to their data. Finally, we advise optimizing the processing resources allocated to Stata when using **quaidsce**, as computation times increase rapidly with the number of categories and observations (both for the estimated model and post-estimation commands). In practical applications, producing elasticity estimates over a nationally representative sample with bootstrap standard errors can take up to several days, using optimized settings on a standard computer.

6. References

Amemiya, T. 1974. Multivariate regression and simultaneous equation models when the dependent variables are truncated normal. *Econometrica* 42(6): 999-1012.

Banks, J., R. Blundell, and A. Lewbel. 1997. Quadratic Engel curves and consumer demand. *The Review of Economics and Statistics* 79: 527-539.

Blundell, R., and J.M. Robin. 1999. Estimation in large and disaggregated demand systems: An estimator for conditionally linear systems. *Journal of Applied Econometrics* 14: 209-232.

Capacci, S., and M. Mazzocchi. 2011. Five-a-day, a price to pay: an evaluation of the UK program impact accounting for market forces. *Journal of Health Economics* 30(1): 87-98.

Crawford, I., Laisney, F., and I. Preston. 2003. Estimation of household demand systems with theoretically compatible Engel curves and unit value specifications. *Journal of Econometrics* 114(2): 221-241.

Deaton, A. 1986. Demand Analysis. In *Handbook of Econometrics*, vol. 3. Ed. Z. Griliches and M.D. Intriligator, 1769-1839 (chapter 30). Amsterdam: North-Holland.

Deaton, A., and J. Muellbauer. 1980. An Almost Ideal Demand System. *American Economic Review* 70(3): 312-326.

Houthakker, H.S. 1953. La forme des courbes d'Engel. *Cahiers du Séminaire d'Econométrie* 2: 59-66.

Instituto Nacional de Estadística de Chile (INE) 2020. VII EPF: Metodología. https://www.ine.cl/estadisticas/sociales/ingresos-y-gastos/encuesta-de-presupuestos-familiares.

Shonkwiler, J.S., and S.T. Yen. 1999. Two-step estimation of a censored system equations. *American Journal of Agricultural Economics* 81: 972-982.

Poi, B.P. 2012. Easy demand-system estimation with quaids. *The Stata Journal* 12(3): 433-446.

Yen, S.T., B.-H. Lin, and D.M. Smallwood. 2003. Quasi and simulated-likelihood approaches to censored demand systems: Food consumption by food stamp recipients in the United States. *American Journal of Agricultural Economics* 85: 458-491.