Censored demand system estimation with *quaidsce*

Juan Carlos Caro University of Luxembourg, Luxembourg

Grace Melo Texas A&M University, US

Jose Alberto Molina, Universidad de Zaragoza, España

Juan Carlos Salgado, Instituto Nacional de Salud Pública, México

Stata Conference 2021, August 6th

What is the issue?

- Zero consumption is very common in household expenditure data.
- Failure to account for censoring can bias demand elasticity estimates, which are key for policy simulation.
- Methods to address censoring in systems of equations have been proposed based on a generalization of Amemiya's model (1974).
- Current Stata user-written commands *quaids* (Poi, 2012) and *aidsills* (Lecocq and Robin, 2015) do not address censoring.
- We proposed *quaidsce*, building upon *quaids* to account for censoring using the two-step estimation approach proposed by Shonkwiler and Yen (1999).

Quadratic Almost Ideal Demand System

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

- where w_{hi} is the budget share of category *i* for household *h and J i*s the number of categories
- p_{hj} is the price index, y_h is total household expenditure; z_{hk} are demand shifters and u_{hi} is the residual.
- Price deflators:

$$a(p_h) = \alpha_o + \sum_{j=1}^{J} \alpha_{ij} \ln p_{hj} + \sum_{j=1}^{J} \sum_{k=1}^{K} \eta_{ik} z_{hk} \ln p_{hj} + 0.5 \sum_{i=1}^{J} \sum_{j=1}^{J} \eta_{ij} \ln p_{hi} \ln p_{hj}$$

$$b(p_h) = exp(\sum_{j=1}^{J} \beta_i \ln p_{hj})$$

- Restrictions on homogeneity and symmetry are imposed

Two-step approach (SY, 1999)

- The latent share w_{hi}^* is related to the observed share as follows:

$$w_{hi} = d_{hi} w_{hi}^*$$

where and d_{hi} is a binary dependent variable that equals 1 for nonzero expenditure.

- The unconditional expected value can be written as:

$$w_{hi}^* = \Phi(x_h'\theta_i)\, w_{hi} + \delta_i\, \phi(x_h{}'\theta_i) + \varepsilon_i$$

where Φ and ϕ are the cumulative and density normal distribution functions.

Two-step approach:

1. a univariate probit equation $d_{hi} = z_h{}'\theta_i \;\forall\; i$ is estimated for all categories, where x_h is a vector of regressors including prices and demographic variables.
2. $\Phi(z_h'\widehat{\theta_i})$ and $\phi\left(z_h'\widehat{\theta_i}\right)$ are calculated and included in the second step as follows:

$$w_{hi}^* = \Phi(z_h'\widehat{\theta_i})\, w_{hi} + \delta_i \phi\left(z_h'\widehat{\theta_i}\right) + \varepsilon_i$$

The system is no longer singular, hence all equations are jointly estimated (no additivity).

quaidsce syntax

- The syntax draws directly from **quaids** with the addition of the censoring option.

```
 quaidsce varlist_expshares [if] [in], anot(#)
{prices(varlist_prices)|lnprices(varlist_lnprices)}
{expenditure(varlist_exp)|lnexpenditure(varlist_lnexp)}

[demographics(varlist_demo) noquadratic nolog nocensor vce(vcetype)]
```

- Additional features (in progress):
  - Automatically correct for expenditure endogeneity using IV
  - Incorporate bootstrapped estimates for elasticities (instead of delta method)
  - Manually adjust *nlsur* estimation method for (testing purposes)

An application

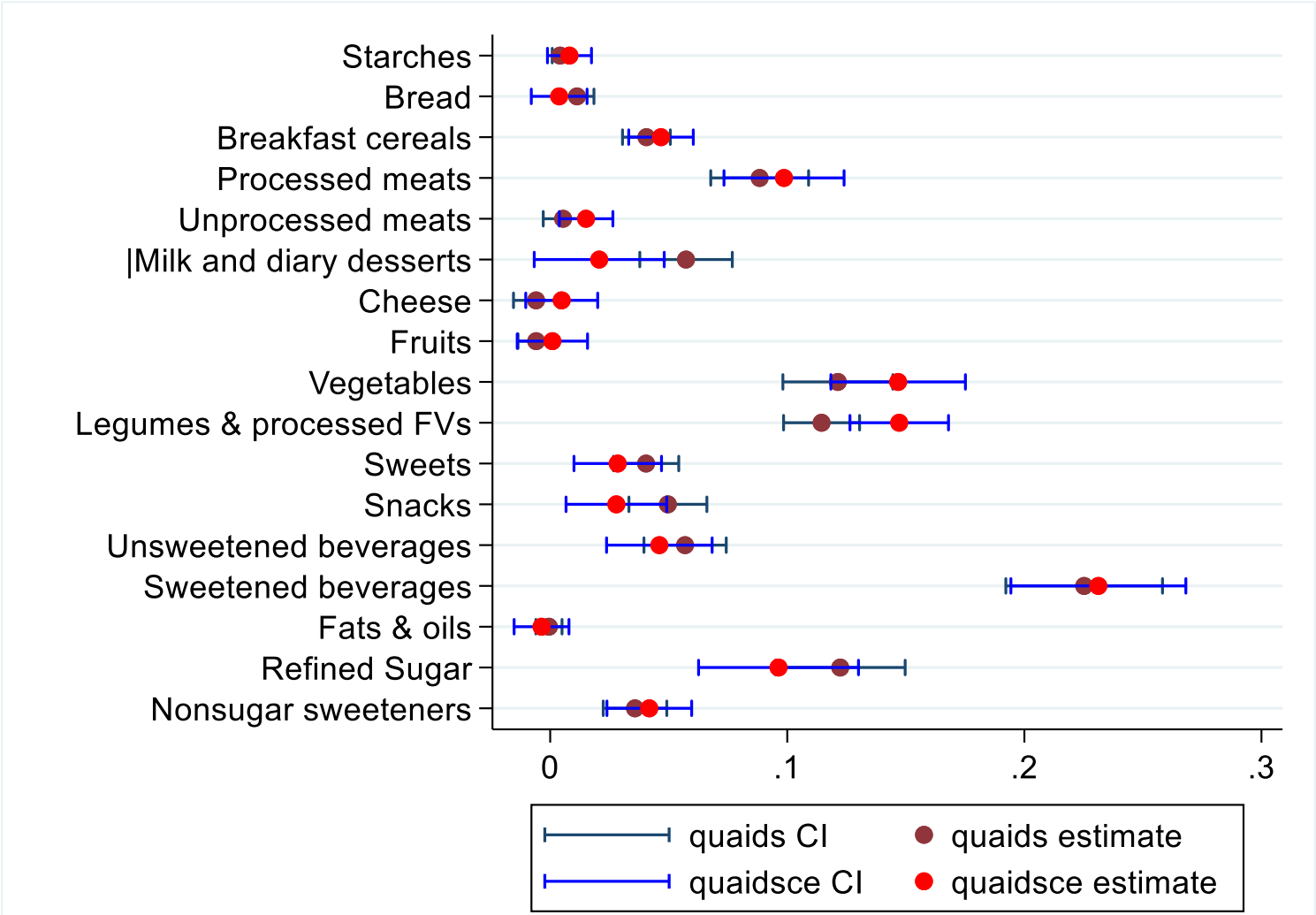
- Demand system estimation for food consumption using 17 categories as in Melo (*forthcoming*).
- Cross-sectional data from the Chilean National Household Budget Survey for 2016/2017 (latest study period).
- Estimates for the demand system compared with *quaids* (Poi, 2012).
- Probit regressions based on prices (quality-adjusted unit values) and randomly generated household covariates.

Data (N=14,703 households)

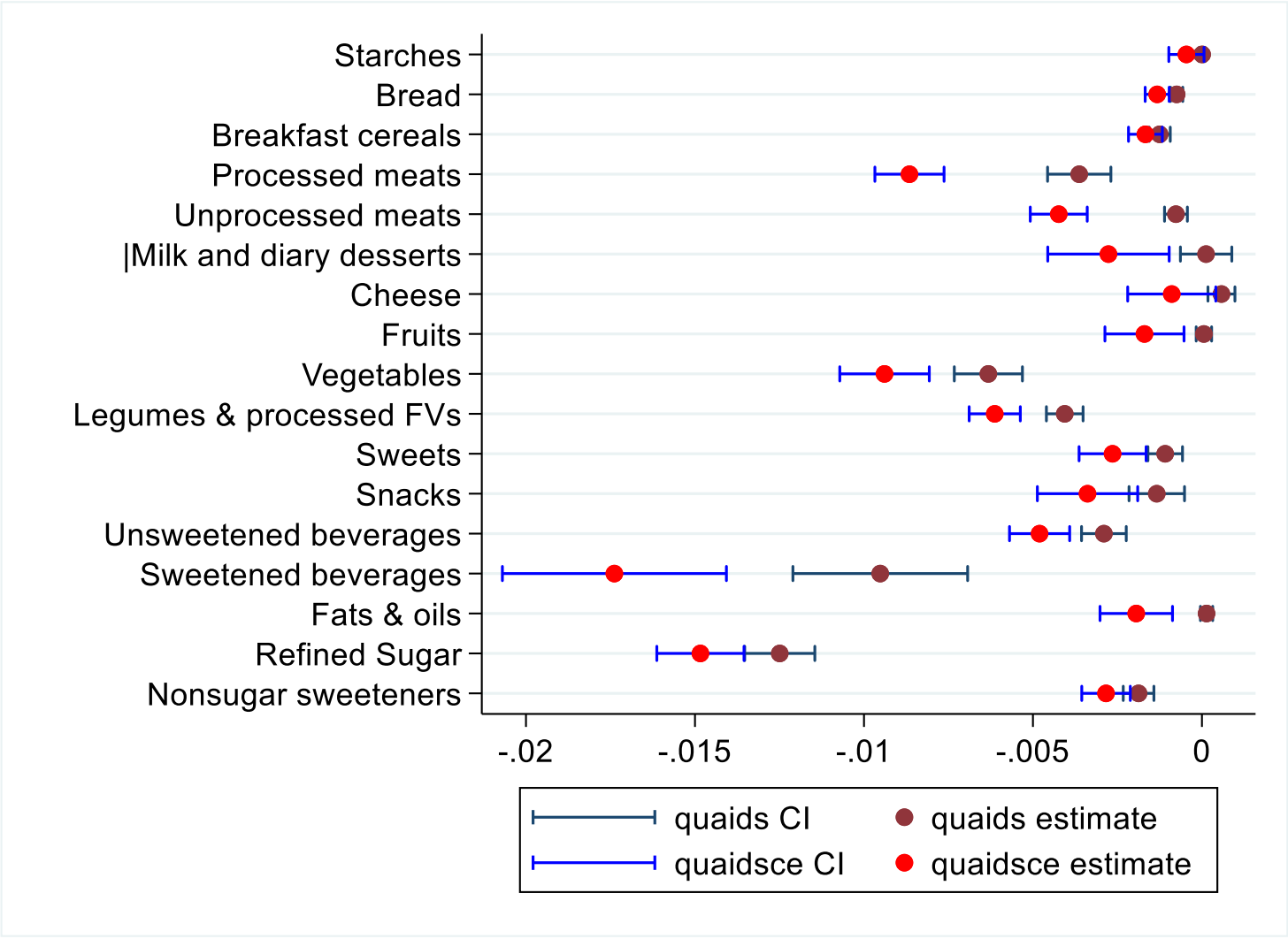
Food group	Purchase > 0	Quantity (g) | Purchase > 0
1 Starches	0.64	89.63
2 Bread	0.97	197.82
3 Breakfast cereals	0.27	20.25
4 Unprocessed meats	0.89	146.69
5 Processed meats	0.83	40.89
6 Milk & dairy desserts	0.74	164.23
7 Cheese	0.72	25.79
8 Fruits	0.78	245.62
9 Vegetables	0.89	212.15
10 Legumes & processed FVs	0.40	24.07
11 Sweets	0.61	36.81
12 Snacks	0.76	38.30
13 Unsweetened beverages	0.58	882.32
14 Sweetened beverages	0.86	287.12
15 Fats & oils	0.64	43.73
16 Refined Sugar	0.33	63.31
17 Nonsugar sweeteners	0.07	7.93

Results: delta coefficients (95% CI)

Results: alpha coefficients (95% CI)



Results: beta coefficients (95% CI)



Summary

- *quaidsce* extends over *quaids* to allow for censoring in consumption and expenditure data, using a two-step approach.
- A practical application highlights the differences between both approaches, using data from the Chilean National Household Budget Survey 2016/2017.
- Due to the non-linear approach and plug-in estimator, standard errors should be estimated via bootstrap.
- Next steps
  - Elasticities and postestimation
  - Bootstrap SE
  - Replications

Censored demand system estimation with *quaidsce*

Juan Carlos Caro University of Luxembourg, Luxembourg

Grace Melo Texas A&M University, US

Jose Alberto Molina, Universidad de Zaragoza, España

Juan Carlos Salgado, Instituto Nacional de Salud Pública, México

Stata Conference 2021, August 6th