Demand system estimation with Stata: Multivariate censoring and other econometric issues

Soufiane KHOUDMI (University of Montpellier/LAMETA)
Benoit MULKAY (University of Montpellier)

Stata Conference New Orleans 2013

Motivation

- In the heart of many economics applied work.
- Deaton & Muellbauer (1980) , Pudney (1985) , Banks, Blundell & Lewbel (1997)
- Welfare, Taxation, Industrial Organisation, Health.
- Hot Topic in `statalist`.
- Poi (2002, 2008, 2012)

In This Presentation...

Focus on the Censoring Correction proposed by Tauchmann (2010)

The procedure violate Invariance (with respect to the dropped equation to avoid singularity) and Adding-Up

Retrieve a unique set of estimator and restricted to adding-up (among other restrictions) with Minimum Distance Estimator following Pudney (1985)

Not a program/command

Quadratic Almost Ideal Demand System Specification

For each good $\ell = \{1..\ell..L\}$, Banks, Blundell and Lewbel's (1997) specifies consumption share w_ℓ ($\sum_\ell w_\ell = 1$) with prices $\mathbf{p} = \{p_1...p_\ell...p_L\}$ and total expenditure m ($\neq$ income) :

$$w_\ell = \alpha_\ell + \sum_j \gamma_{\ell j} \ln p_j + \beta_\ell \ln\left[\frac{m}{a(\mathbf{p})}\right] + \frac{\lambda_\ell}{b(\mathbf{p})}\left\{\ln\left[\frac{m}{a(\mathbf{p}}\right]\right\}^2$$

With the deflators/agregators :

Translog : $\ln a(\mathbf{p}) = \alpha_0 + \sum_\ell \alpha_\ell \ln p_\ell + \frac{1}{2}\sum_\ell \sum_j \gamma_{\ell j} \ln p_\ell \ln p_j$

Cobb Douglas : $b(\mathbf{p}) = \prod_\ell p_\ell^{\beta_\ell}$

Satifiying the behavioral assumptions restrictions :

Adding Up : $\sum_\ell \alpha_\ell = 1$, $\beta_\ell = 0$, $\sum_\ell \lambda_\ell = 0$ and $\sum_\ell \gamma_{\ell j} = 0 \, \forall j$

Homogeneity : $\sum_j \gamma_{\ell j} = 0 \, \forall \ell$

Symmetry : $\gamma_{\ell j} = \gamma_{j\ell} \, \forall \ell, j$

Sample Description

	Mean	S.E.	Median	%	Comments
1 Food	0,318	0,158	0,302	0,99	Food and Drinks
2 Housing	0,189	0,135	0,153	0,99	Energy Water Utilities (No Rent)
3 Outside	0,279	0,188	0,268	0,91	Transport Recreational
4 Alcohol	0,027	0,049	0,008	0,60	Wine Beer and Liquors
4' Alcohol	0,045	0,057	0,027		If consummed , $N = 6099$
5 Clothing	0,130	0,122	0,100	0,87	Including Reparation
5'Clothing	0,150	0,120	0,121		If consummed , $N = 8909$
6 Health	0,058	0,099	0,015	0,63	Out Of Pocket
6' Health	0,091	0,112	0,053		If consummed , $N = 6501$
Expenditure (€)	14856	10365	12441		$\approx$ 50 % of overall reported

From Budget Des Familles (BDF) 2005/2006 : Expenditure Famillies Survey
Sample Household 10240 No survey weighting
S.E. stands for Standard Errors. % stands for selection.

Setup : Type II Tobit

Amemiya (1985) classification. For each equation ℓ :

Latent (*)

Selection : $d_\ell^* = z'\eta + \nu_\ell$

Result : $w_\ell^* = x'\theta + \epsilon_\ell$

Observed

Selection $d_\ell = 1$ if $w_\ell > 0$

Selection $d_\ell = 0$ otherwise

Result $w_\ell = d_\ell w_\ell^*$

Where :

$$\begin{pmatrix} \nu \\ \epsilon \end{pmatrix} \sim 2L\ Normal \left[0; \begin{pmatrix} \Sigma_{\nu\nu} & \Sigma'_{\nu\epsilon} \\ \Sigma_{\nu\epsilon} & \Sigma_{\epsilon\epsilon} \end{pmatrix} \right]$$

Simple Case - 1 equation

Heckman (1979) : "Mispecification" Problem :

$$\begin{aligned} E(w_\ell | d_\ell > 0) &= x'\theta_\ell + E(\epsilon_\ell | d_\ell > 0) \\ &= x'\theta_\ell + \theta_\ell^C E(\nu_\ell | \nu_\ell > -z'\eta_\ell) \\ &= x'\theta_\ell + \theta_\ell^C \frac{\phi(z'\eta_\ell)}{\Phi(z'\eta_\ell)} \\ &= x'\theta_\ell + \theta_\ell^C IMR_\ell \end{aligned}$$

Two-step estimator :

- (i) Compute the Inverse Mills ratio (prediction's p.d.f on c.d.f) after a Probit on selection equation.
- (ii) Then estimate the augmented result equation on the observed part.

In Stata, it may be `heckman`, or :

(i)
```
probit d z1 z2
predict zeta
generate IMR = normalden(zeta)/normal(zeta)
```
(ii)
```
regress w x1 x2 IMR if w>0
```

General Case - L equations

Tauchmann (2010) conditionnate on $\mathbf{d} = \{d_1 ... d_\ell ... d_L\}$:

$$E(w_\ell|\mathbf{d}) = x'\theta_\ell + E(\epsilon_\ell|\mathbf{d}) = x'\theta_\ell + \sum_{j=1}^{L} \sigma_{\epsilon\ell,\nu j} E(\nu_j|\mathbf{d})$$

Using Tallis (1961), the system to estimate :

$$w_\ell = d_\ell x'\theta_\ell + d_\ell \sum_j \theta^c_{\ell j} \psi_j \phi(\psi_j z'\eta_j) \Phi^{(L-1)}(A_j, \Psi_j R_j \Psi_j)/\Phi(\mathbf{d})$$

$$w_\ell = d_\ell x'\theta_\ell + d_\ell \sum_j \theta^c_{\ell j} TT\text{-}IMR_j$$

With :

- $\Phi^{(L-1)}$ is $(L-1)$ dimentionnal c.d.f
- $\Phi^{(L)}(\mathbf{d})$ is probability of the observed pattern
- $\psi_j = 2d_j - 1$ such that : $d_j = \{0;1\} \rightarrow \psi_j = \{-1;1\}$
- Ψ_j is a ψ_j elements diagonal matrix
- $A_j = \psi_j(z'\eta_j - \sigma_{\nu i,\nu j} z'\eta_i)/(1 - (\sigma^2_{\nu i,\nu j})^{1/2})\ i \neq j$
- R_j is partial correlation $(L-1)$ matrix $Cor(\nu|\nu_j)$

Data Suggestion

j	ℓ	ℓ Mean if j not selected	ℓ Mean if j selected	t- statistic
Alcohol	Food	0,3028	0,3275	-7,52
	Clothing	0,1404	0,1235	6,65
	Clothing*	0,1662	0,1392	10,13
	Housing	0,2143	0,1714	15,08
	Health	0,0598	0,0568	1,48
	Health*	0,1009	0,0856	5,18
	Outside	0,2827	0,2757	1,78

Using `ttest` , * When ℓ and j both are selected.
$\chi^2(1)$: Alcohol vs Clothing (39) ; Alcohol vs Health (52) Clothing vs Health (169)

Comments...

$$\begin{aligned} d_\ell^* &= z'\eta + \nu_\ell \\ w_\ell^* &= x'\theta + \epsilon_\ell \end{aligned} \text{ with } \begin{pmatrix} \nu \\ \epsilon \end{pmatrix} \sim 2L\, Normal \left[0; \begin{pmatrix} \Sigma_{\nu\nu} & \Sigma'_{\nu\epsilon} \\ \Sigma_{\nu\epsilon} & \Sigma_{\epsilon\epsilon} \end{pmatrix} \right]$$

- Some Intermediate Cases

$\Sigma_{\nu\nu}$	$\Sigma_{\nu\epsilon}$	$\Sigma_{\epsilon\epsilon}$	Then
Diag	Diag	Diag	OLS : $w_\ell = d_\ell x'\theta_\ell + d_\ell \theta_\ell^c IMR_\ell$
Diag	Diag	Dense	SUR : $w_\ell = d_\ell x'\theta_\ell + d_\ell \theta_\ell^c IMR_\ell$
Diag	Dense	Dense	SUR : $w_\ell = d_\ell x'\theta_\ell + d_\ell \sum_j \theta_{\ell j}^c IMR_j$
Dense	Dense	Dense	SUR : $w_\ell = d_\ell x'\theta_\ell + d_\ell \sum_j \theta_{\ell j}^c TT\text{-}IMR_j$

Issues :

× Unconsistency of estimators variance-covariance matrix Σ_θ.
Not so innocent issue since we use GLS and Minimum Distance.

- The demand system is Invariant and Adding Up is no longer Garanted.

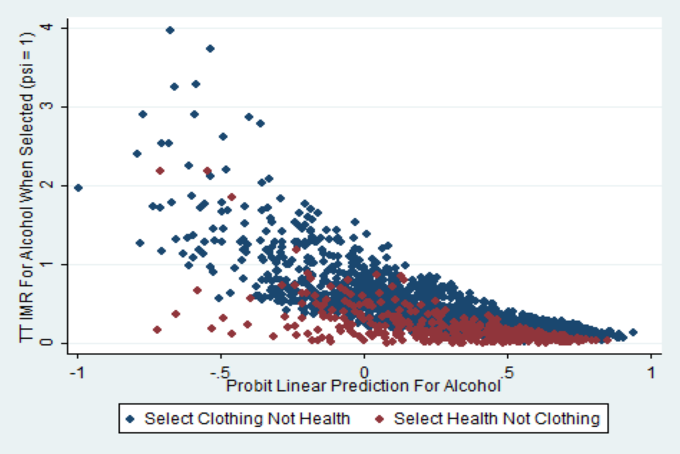
Multivariate Probit

Equation →	Alcohol		Clothing		Health	
Allocation	-0.098	**	0.371	***	0.12	**
Chidrens Number	-0.063	*	0.228	***	0.128	***
HH Head Age	0.048	***	0.017	***	0.012	**
Adults Number	0.561	***	0.776	***	0.612	***
HH Head Male	0.108	***	-0.367	***	-0.205	***
HH Head French	0.379	***	0.024		0.198	***
$\sigma_{\nu \mathbf{Alc}, \nu \mathbf{Cloth}}$	0.072	***				
$\sigma_{\nu \mathbf{Alc}, \nu \mathbf{Cloth}}$	0.072	***				
$\sigma_{\nu \mathbf{Alc}, \nu \mathbf{Healt}}$	0.047	***				
$\sigma_{\nu \mathbf{Cloth}, \nu \mathbf{Health}}$	0.165	***				

Multivariate Probit using Cappellari and Jenkins `mvnp`
(An Easy to go alternative for 3 equations : Terracol (2002) `triprobit` program
Income, Localisation and Squares Variables Output Omited
*** , ** ,* resp for 1% , 5 % and 10% signficativity.

Multivariate Inverse Mills Ratio

Recall That $TT\text{-}IMR_j$: $\psi_j \phi(\psi_j z' \eta_j) \Phi^{(L-1)}(A_j, \Psi_j R_j \Psi_j) / \Phi^L(\mathbf{d})$
With $\hat{\eta}$ and $\Sigma_{\nu\nu}$, we get c.d.f. using `mvnp` or `binormal` when $L = 2$.



Estimation Strategy

Poi (2008) NLSUR proposes a framework for a "complete" estimation with `NLSUR`.

We also follow Banks & al. (1997), Blundell (1988) , Blundell & Robin (1999)

- Impose Homogeneity expressing all prices relative to a commodity price.
- Impose Symmetry using Minimum Distance Estimator.
  This last is used to impose Adding Up.
- Iterative Procedure to Allow For non Linearity.
- Test and Allow for Expenditure Endogeneity with a Durbin Wu Hausmann term
  The residual from a first stage regression : expenditure on instruments (with income)

Estimation Results

The system is sensible to the dropped equation
The constant α_ℓ and expenditure associated parameter β_ℓ for each equation ℓ and with respect to each excluded equation :

excluded $\rightarrow$	food	alcohol	clothing	housing	health	outside
α food	0.000	0.057	0.077	0.076	0.074	0.071
s.e.	0.000	0.042	0.042	0.042	0.042	0.042
α alcohol	0.133	0.000	0.146	0.138	0.145	0.152
	0.024	0.000	0.024	0.024	0.024	0.025
α clothing	0.232	0.127	0.000	0.120	0.131	0.210
	0.044	0.032	0.000	0.042	0.039	0.046
α housing	0.587	0.578	0.589	0.000	0.589	0.594
	0.036	0.035	0.036	0.000	0.036	0.036
α health	0.385	0.325	0.354	0.356	0.000	0.330
	0.045	0.033	0.043	0.044	0.000	0.048
α outside	0.153	0.126	0.162	0.183	0.163	0.000
	0.048	0.047	0.049	0.048	0.049	0.000
β food	0.000	0.106	0.094	0.096	0.098	0.099
	0.000	0.017	0.017	0.017	0.017	0.017
β alcohol	-0.029	0.000	-0.031	-0.028	-0.031	-0.035
	0.010	0.000	0.009	0.010	0.009	0.010
β clothing	-0.034	-0.009	0.000	0.011	0.002	-0.021
	0.018	0.013	0.000	0.017	0.016	0.019
β housing	-0.116	-0.111	-0.118	0.000	-0.117	-0.120
	0.015	0.015	0.015	0.000	0.015	0.015
β health	-0.122	-0.096	-0.102	-0.107	0.000	-0.098
	0.018	0.013	0.017	0.017	0.000	0.019
β outside	0.041	0.054	0.034	0.027	0.039	0.000
	0.020	0.019	0.020	0.020	0.020	0.000

TT-IMR

$\theta^c_{\ell j}$ (TT-IMR associated parameter) to assess the selectivity effect on each equation.

TT-IMR $\rightarrow$	Alcohol	Cloth	Health
Food	+	-	-
Alcohol	-	-	-
Clothing	-	$\times$	-
Housing	-	-	-
Health	-	-	$\times$ or +
Outside	-	-	-

- Means affect negatively , + means affect positively , $\times$ means ambigueous effet.

More precisely : $\theta^c_{\ell j} \times TT\text{-}IMR_j$ to interpret this. To be evaluated at the average.

MDE : Unique Set Of Parameters

We follow Greene (2012) with a mapping matrix $K_{-\ell}$:

$$\text{minimize}_{\theta^U}\ C =$$

$$\begin{pmatrix} (K_{-1}\hat{\theta}_{-1} - K_{-1}\theta^U) \\ \vdots \\ (K_{-\ell}\hat{\theta}_{-\ell} - K_{-\ell}\theta^U) \\ \vdots \\ (K_{-L}\hat{\theta}_{-L} - K_{-L}\theta^U) \end{pmatrix}' \begin{pmatrix} \Sigma_{-1} & 0 & \dots & 0 \\ 0 & \Sigma_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Sigma_{-L} \end{pmatrix}^{-1} \begin{pmatrix} (K_{-1}\hat{\theta}_{-1} - K_{-1}\theta^U) \\ \vdots \\ (K_{-\ell}\hat{\theta}_{-\ell} - K_{-\ell}\theta^U) \\ \vdots \\ (K_{-L}\hat{\theta}_{-L} - K_{-L}\theta^U) \end{pmatrix}$$

Exemple : L=3 , K_{-3} is given by :

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{b}_3 \end{pmatrix} - \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} b_1^U \\ b_2^U \\ b_3^U \end{pmatrix} = \begin{pmatrix} \hat{b}_1 \\ \hat{b}_2 \end{pmatrix} - \begin{pmatrix} b_1^U \\ b_2^U \end{pmatrix}$$

MDE : Unique Set Of Parameters

Using `optimize` :

excluded	food	alcohol	clothing	housing	health	outside	average	MD Uniq	Sum
α food	0.000	0.057	0.077	0.076	0.074	0.071	0.071	0.078	
s.e.	0.000	0.042	0.042	0.042	0.042	0.042		0.012	
α alcohol	0.133	0.000	0.146	0.138	0.145	0.152	0.143	0.136	
	0.024	0.000	0.024	0.024	0.024	0.025		0.007	
α clothing	0.232	0.127	0.000	0.120	0.131	0.210	0.164	0.128	
	0.044	0.032	0.000	0.042	0.039	0.046		0.012	
α housing	0.587	0.578	0.589	0.000	0.589	0.594	0.587	0.550	
	0.036	0.035	0.036	0.000	0.036	0.036		0.010	
α health	0.385	0.325	0.354	0.356	0.000	0.330	0.350	0.311	
	0.045	0.033	0.043	0.044	0.000	0.048		0.012	
α outside	0.153	0.126	0.162	0.183	0.163	0.000	0.157	0.183	1.39
	0.048	0.047	0.049	0.048	0.049	0.000		0.014	
β food	0.000	0.106	0.094	0.096	0.098	0.099	0.098	0.096	
	0.000	0.017	0.017	0.017	0.017	0.017		0.005	
β alcohol	-0.029	0.000	-0.031	-0.028	-0.031	-0.035	-0.031	-0.028	
	0.010	0.000	0.009	0.010	0.009	0.010		0.003	
β clothing	-0.034	-0.009	0.000	0.011	0.002	-0.021	-0.010	0.003	
	0.018	0.013	0.000	0.017	0.016	0.019		0.005	
β housing	-0.116	-0.111	-0.118	0.000	-0.117	-0.120	-0.116	-0.101	
	0.015	0.015	0.015	0.000	0.015	0.015		0.004	
β health	-0.122	-0.096	-0.102	-0.107	0.000	-0.098	-0.105	-0.091	
	0.018	0.013	0.017	0.017	0.000	0.019		0.005	
β outside	0.041	0.054	0.034	0.027	0.039	0.000	0.039	0.029	-0.09
	0.020	0.019	0.020	0.020	0.020	0.000		0.006	

MDE : Restricted Parameters

Same than previsously :
With $\hat{\theta}_U$ and $\hat{\Sigma}_U$ in hand, we want θ_R :

$$\text{minimize}_{\theta_R} \; C = (\hat{\theta}_U - K\theta_R)'\hat{\Sigma}_U^{-1}(\hat{\theta}_U - K\theta_R)$$

Ex. with symmetry and $L = 2$ for γ's matrix K bloc such as :

$$\begin{pmatrix} \gamma_{11}^U \\ \gamma_{12}^U \\ \gamma_{21}^U \\ \gamma_{22}^U \end{pmatrix} - \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} \gamma_{11}^R \\ \gamma_{12}^R \\ \gamma_{22}^R \end{pmatrix}$$

Ex. with adding up and $L = 3$ for β's matrix K bloc such as :

$$\begin{pmatrix} \hat{\beta}_1^U \\ \hat{\beta}_2^U \\ \hat{\beta}_3^U \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix} \cdot \begin{pmatrix} \beta_1^R \\ \beta_2^R \end{pmatrix}$$

MDE : Restricted Parameters

Using optimize :

	MDE Unique	Sum	MDE Restriction
α food	0.078		0.089
s.e.	0.012		0.008
α alcohol	0.136		0.061
	0.007		0.005
α clothing	0.128		0.028
	0.012		0.008
α housing	0.550		0.578
	0.010		0.007
α health	0.311		0.027
	0.012		0.007
α outside	0.183	1.39	
	0.014		
β food	0.096		0.092
	0.005		0.003
β alcohol	-0.028		-0.010
	0.003		0.002
β clothing	0.003		0.038
	0.005		0.003
β housing	-0.101		-0.106
	0.004		0.003
β health	-0.091		-0.024
	0.005		0.003
β outside	0.029	-0.09	
	0.006		

Elasticities

With Tauchmann (2010) multivariate censoring correction :

	Exp.	Food P.	Alcohol P.	Clothing P.	Housing P.	Health P.	Outside P.
Food	0.541	**-0.263**	-0.210	-0.082	0.061	-0.112	0.066
Alcohol	0.981	-0.469	**-1.452**	0.438	0.228	0.110	0.165
Clothing	1.203	-0.218	0.213	**-1.353**	-0.026	0.044	0.137
Housing	0.606	0.094	0.164	0.064	**-1.236**	0.201	0.107
Health	1.079	-0.161	0.049	0.038	0.116	**-1.220**	0.098
Outside	1.314	-0.149	0.037	0.071	-0.057	0.062	**-1.278**

Whitout any censoring correction :

	Exp.	Food P.	Alcohol P.	Clothing P.	Housing P.	Health P.	Outside P.
Food	0.549	**-0.272**	-0.189	-0.221	0.245	-0.192	0.081
Alcohol	0.777	-0.269	**-1.355**	0.269	0.173	0.142	0.264
Clothing	1.046	-0.222	0.145	**-1.413**	-0.009	0.009	0.443
Housing	0.599	0.352	0.065	-0.026	**-1.019**	0.207	-0.178
Health	1.064	-0.238	0.065	0.015	0.121	**-1.245**	0.219
Outside	1.470	-0.261	0.068	0.298	-0.258	0.134	**-1.451**

In This Presentation ...

- Multivariate Type II Tobit with Heckman (1979) type augmented two-step approach

  Iterative Procedure : Conditionnal Linearity

  Endogeneity with Durbin-Wu-Hausman

  Homogeneity with Relative Prices
- Unique set of estimators with Minimum Distance
- Adding up and symmetry restricted parameters with Minimum Distance

Some points are necessary for the censored demand system estimation,
Some points are alternatives or complement to Poi (2008).

Thank You !

khoudmi@lameta.univ-montp1.fr