Econ 673: Microeconometrics

Chapter 8: Limited Dependant Variable Models

Outline

- Truncation
  - Direct
  - Incidental
- Censoring
  - Univariate
  - Multivariate
    - Amemiya-Tobin
    - Kuhn-Tucker

Truncation Versus Censoring

- Differ in term of the degree of missing data
- Consider a latent variable of interest modeled as a linear function of observable characteristics

$$y_i^* = \beta' x_i + \varepsilon_i$$

- Truncation: Observe y_i^* s.t. $y_i^* \in A$; e.g., $A = [a,b]$

$$y_i = \begin{cases} y_i^* & y_i^* \in A \\ \text{no observations} & y_i^* \notin A \end{cases}$$

- Censoring: Observe

$$y_i = \begin{cases} y_i^* & y_i^* \in A \\ h\left(y_i^*\right) & y_i^* \notin A \end{cases}$$

Examples

- Truncation
  - Low income household studies (e.g., Hausman and Wise, 1977)
  - On-site visitation data
  - TOU pricing experiments – excluded low-usage households
  - Employment data on hours worked (excludes unemployed)
- Censoring
  - Capacity constrained data (e.g., class enrollments or ticket sales)
  - Hours worked (or leisure demand) – essentially capacity constrained
  - Commodity purchases (non-negative)

Illustration

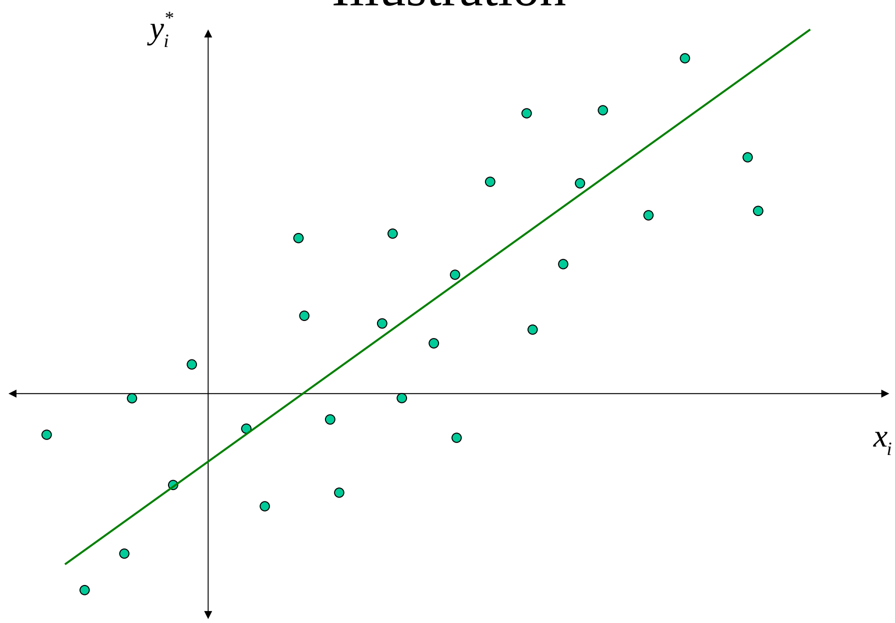


Illustration - Truncation

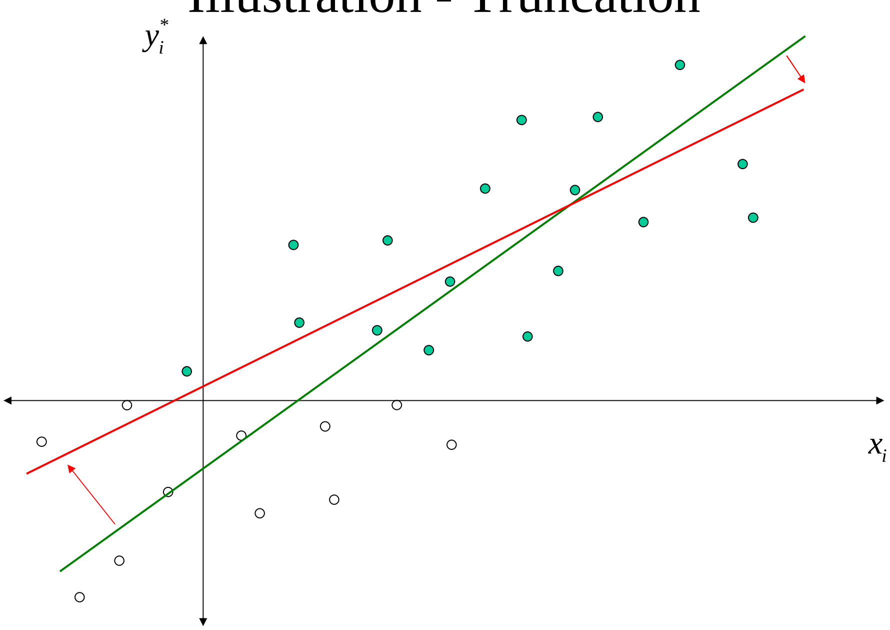
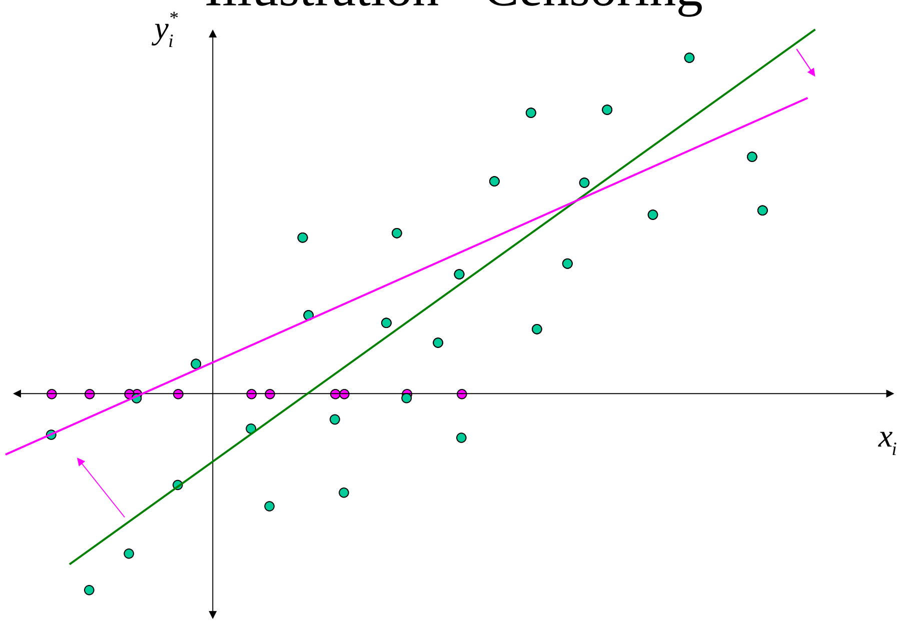


Illustration - Censoring



Sources - Truncation

- *Greene, W. H., (2000) *Econometric Analysis*, 4th edition, Upper Saddle River, New Jersey: Prentice-Hall, Inc., Sections 20.1, 20.2, and 20.4.
- Ruud, P., (2000) *An Introduction to Classical Econometric Theory*, New York: Oxford University Press, Ch. 28.
- Maddala, G. S., (1983) *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge: Cambridge University Press, Ch. 6.
- Mittelhammer, R. C., G. G. Judge, and D. J. Miller (2000), *Econometric Foundations*, Cambridge, MA: Cambridge University Press, Section 20.4.

Truncation

- The fundamental problem is that truncation will alter the distributional characteristics of the available sample
- Consider linear regression model

$$y_i^* = \beta' x_i + \varepsilon_i$$

where

$$E\left[\varepsilon_i \mid x_i\right] = 0$$

and

$$Var\left[\varepsilon_i \mid x_i\right] = \sigma^2$$

so that

$$E\left[y_i \mid x_i\right] = \beta' x_i$$

These moments will typically not apply for the truncated sample

Consider a Truncation from Below

$$\begin{aligned} E\left[y_i \mid x_i, y_i^* \geq a_L\right] &= \beta' x_i + E\left[\varepsilon_i \mid x_i, y_i^* \geq a_L\right] \\ &= \beta' x_i + E\left[\varepsilon_i \mid x_i, \beta' x_i + \varepsilon_i \geq a_L\right] \\ &= \beta' x_i + E\left[\varepsilon_i \mid \varepsilon_i \geq a_L - \beta' x_i\right] \\ &= \beta' x_i + \int_{a_L - \beta' x_i}^{\infty} \varepsilon f\left(\varepsilon \mid \varepsilon \geq a_L - \beta' x_i\right) d\varepsilon \\ &= \beta' x_i + \int_{\alpha_{Li}}^{\infty} \varepsilon f\left(\varepsilon \mid \varepsilon \geq \alpha_{Li}\right) d\varepsilon \quad > \beta' x_i \end{aligned}$$

where

$$f(\varepsilon_i \mid \varepsilon_i \geq \alpha) = \frac{f(\varepsilon_i)}{\Pr\left[\varepsilon_i \geq \alpha\right]} = \frac{f(\varepsilon_i)}{1 - F(\alpha)}$$

and

$$\alpha_{Li} \equiv a_L - \beta' x_i$$

Truncation from Below Assuming Normality

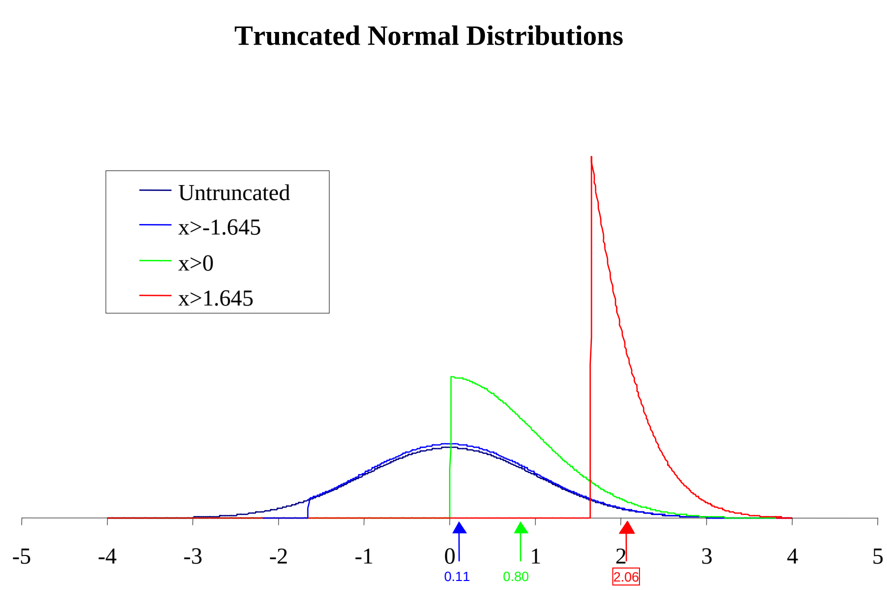
If

$$\varepsilon_i \sim N\left(0, \sigma^2\right)$$

$$f(\varepsilon_i \mid \varepsilon_i \geq \alpha_{Li}) = \frac{\frac{1}{\sigma}\phi\left(\frac{\varepsilon_i}{\sigma}\right)}{1 - \Phi\left(\frac{\alpha_{Li}}{\sigma}\right)}$$

so that

$$f(y_i \mid y_i^* \geq a_L) = \frac{\frac{1}{\sigma}\phi\left(\frac{y_i - \beta' x_i}{\sigma}\right)}{1 - \Phi\left(\frac{a_L - \beta' x_i}{\sigma}\right)}$$



Truncation from Below Assuming Normality (cont'd)

Using Theorem 20.2 in Greene (2000, p. 899),

$$E\left[y_i \mid y_i^* \geq a_L\right] = \beta' x_i + E\left[\varepsilon_i \mid \varepsilon_i \geq \alpha_{Li}\right]$$
$$= \beta' x_i + \sigma\lambda\left(\frac{\alpha_{Li}}{\sigma}\right)$$

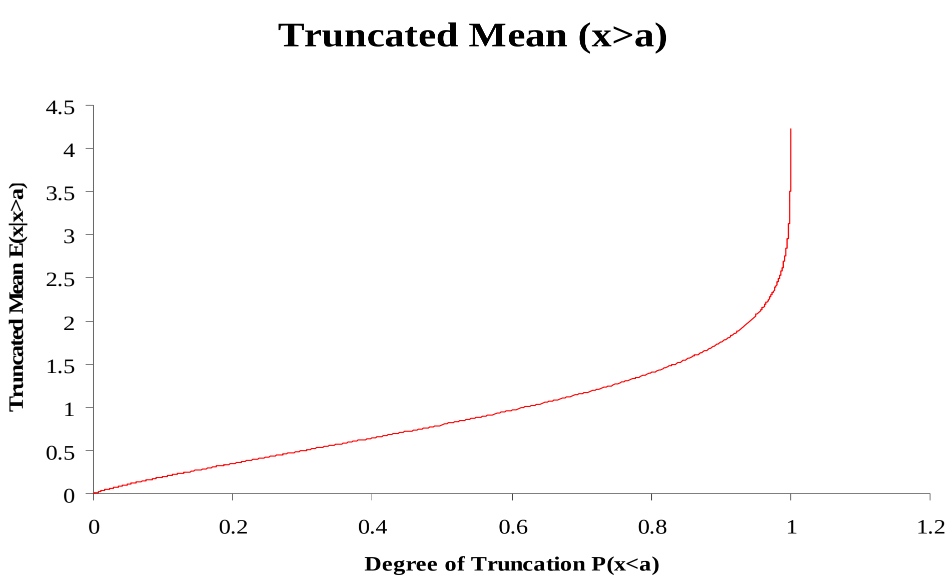
$$\lambda(z) \equiv \frac{\phi(z)}{1-\Phi(z)}$$ Inverse Mill's ratio

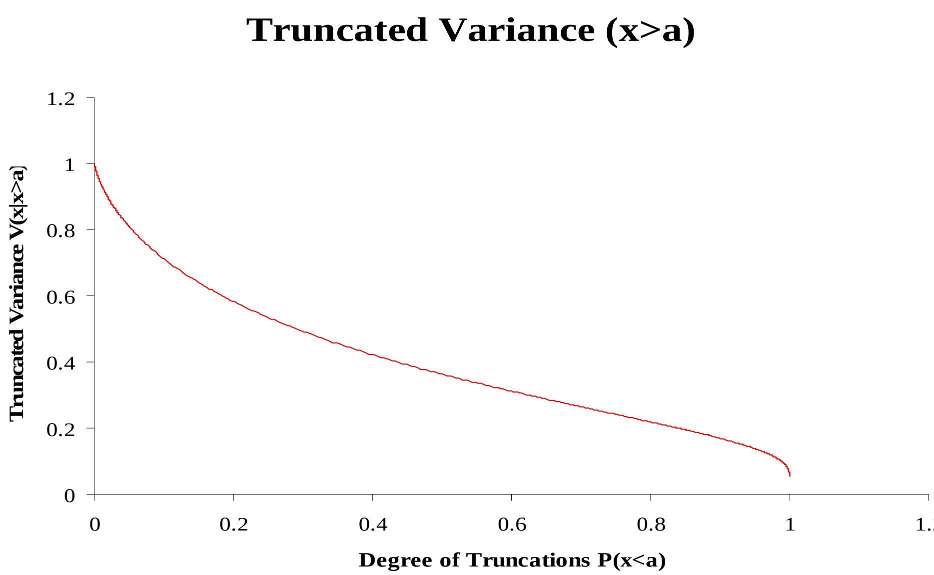
Also

$$V\left[\varepsilon_i \mid y_i^* \geq a_L\right] = \sigma^2\left[1-\delta\left(\frac{\alpha_{Li}}{\sigma}\right)\right]$$

where

$$\delta(z) = \lambda(z)\left[\lambda(z)-z\right] \in (0,1)$$





More General Truncation

- Suppose that the truncation takes the form

$$y_i^* \in B_i = \left[a_{Li}, a_{Ui} \right]$$

$$\Rightarrow$$

$$\varepsilon_i \in \tilde{B}_i = \left[a_{Li} - \beta' x_i, a_{Ui} - \beta' x_i \right] = \left[\alpha_{Li}, \alpha_{Ui} \right]$$

$$f\left(\varepsilon_i \mid \varepsilon_i \in \left[\alpha_{Li}, \alpha_{Ui}\right]\right) = \frac{\frac{1}{\sigma}\phi\left(\frac{\varepsilon_i}{\sigma}\right)}{\left[\Phi\left(\frac{\alpha_{Ui}}{\sigma}\right) - \Phi\left(\frac{\alpha_{Li}}{\sigma}\right)\right]}$$

More General Truncation (cont'd)

$$E\left[y_i \mid y_i^* \in B_i\right] = \beta' x_i + E\left[\varepsilon_i \mid \varepsilon_i \in \tilde{B}_i\right]$$
$$= \beta' x_i + \sigma\lambda\left(\frac{\alpha_{Li}}{\sigma}, \frac{\alpha_{Ui}}{\sigma}\right)$$

where

$$\lambda(z_L, z_U) \equiv \frac{\phi(z_L) - \phi(z_U)}{\Phi(z_U) - \Phi(z_L)}$$

Also

$$V\left[\varepsilon_i \mid \varepsilon_i \in \tilde{B}_i\right] = \sigma^2\left[1 - \delta\left(\frac{\alpha_{Li}}{\sigma}, \frac{\alpha_{Ui}}{\sigma}\right)\right]$$

$$0 < \delta(z_L, z_U) < 1, \quad \delta_{z_L} > 0 \text{ and } \delta_{z_U} < 0$$

Estimation

- OLS estimator of β will be biased, since

$$E\left[b \mid x_i, y_i^* \in B_i\right] = \beta + E\left[(X'X)^{-1} X'\varepsilon \mid X, y_i \in B_i; i = 1, \ldots, N\right]$$
$$= \beta + (X'X)^{-1} X' E\left[\varepsilon \mid \varepsilon_i \in \tilde{B}_i; i = 1, \ldots, N\right]$$
$$\neq \beta$$

- NLS can be used, solving

$$\left(\hat{\beta}_{NLS}, \hat{\sigma}_{NLS}\right) = \arg\min_{\beta,\sigma}\left\{\left(y_i - E\left[y_i \mid y_i^* \in B_i\right]\right)^2\right\}$$
$$= \arg\min_{\beta,\sigma}\left\{\left(y_i - \left[\beta' x_i + \sigma\lambda\left(\frac{\alpha_{Li}}{\sigma}, \frac{\alpha_{Ui}}{\sigma}\right)\right]\right)^2\right\}$$

Estimation (cont'd)

- NLS ignores the heteroskedasticity
- FWNLS can be used noting that

$$V\left[\varepsilon_i \mid \varepsilon_i \in \tilde{B}_i\right] = \sigma^2\left[1-\delta\left(\frac{\alpha_{Li}}{\sigma},\frac{\alpha_{Ui}}{\sigma}\right)\right]$$

from which on can construct appropriate weighting factor – see Ruud (2000, p. 805, eq. 28.20) or Greene (2000, p. 903)

- Most analysts use ML to estimate coefficients

ML Estimation for the Truncated Sample

- We saw earlier that

$$f\left(\varepsilon_i \mid \varepsilon_i \in \left[\alpha_{Li},\alpha_{Ui}\right]\right) = \frac{\frac{1}{\sigma}\phi\left(\frac{\varepsilon_i}{\sigma}\right)}{\left[\Phi\left(\frac{\alpha_{Ui}}{\sigma}\right)-\Phi\left(\frac{\alpha_{Li}}{\sigma}\right)\right]}$$

- The corresponding log-likelihood function becomes

$$LL(\beta,\sigma) = -\frac{N}{2}\left[\log(2\pi)+\log\left(\sigma^2\right)\right]-\frac{1}{2\sigma^2}\sum_{i=1}^{N}\left(y_i-\beta' x_i\right)^2$$
$$-\sum_{i=1}^{N}\log\left[\Phi\left(\frac{a_{Ui}-\beta' x_i}{\sigma}\right)-\Phi\left(\frac{a_{Li}-\beta' x_i}{\sigma}\right)\right]$$

ML Estimation for the Truncated Sample Reparameterization

- It is often convenient to use Olsen's (1976) reparameterization

$$\theta = \sigma^{-1}$$
$$\gamma = \sigma^{-1}\beta$$

- The log-likelihood function then becomes

$$LL(\gamma,\theta) = -\frac{N}{2}\left[\log(2\pi) - 2\log(\theta)\right] - \frac{1}{2}\sum_{i=1}^{N}\left(\theta y_i - \gamma' x_i\right)^2$$
$$-\sum_{i=1}^{N}\log\left[\Phi\left(\theta a_{Hi} - \gamma' x_i\right) - \Phi\left(\theta a_{Li} - \gamma' x_i\right)\right]$$

ML Estimation for the Truncated Sample Reparameterization (cont'd)

- Consider the special case, with

$$a_{Hi} = \infty, a_{Li} = 0$$

- The log-likelihood function becomes

$$LL(\gamma,\theta) = -\frac{N}{2}\left[\log(2\pi) - 2\log(\theta)\right] - \frac{1}{2}\sum_{i=1}^{N}\left(\theta y_i - \gamma' x_i\right)^2 - \sum_{i=1}^{N}\log\left[\Phi\left(\gamma' x_i\right)\right]$$

$$LL_\gamma = \sum_{i=1}^{N}\left(y_i - \gamma' x_i - \lambda_i\right)x_i \qquad LL_\theta = \sum_{i=1}^{N}\left[\frac{1}{\theta} - \left(y_i - \gamma' x_i\right)\lambda_i\right]$$

Hessian can be readily computed for use in estimation

ML (cont'd)

- It has not been shown that the log-likelihood function is globally concave
- Orme and Ruud (1998) have, however, shown that there is a unique MLE

Interpretation

- For the general population

$$y_i^* = \beta' x_i + \varepsilon_i$$

and

$$\frac{\partial E\left(y_i^*\right)}{\partial x} = \beta$$

- However, for the sub-population represented by the truncated sample

$$\frac{\partial E\left[y_i \mid y_i^* \in B_i\right]}{\partial x} = \beta + \sigma \frac{\partial}{\partial x}\left[\lambda\left(\frac{\alpha_{Li}}{\sigma}, \frac{\alpha_{Ui}}{\sigma}\right)\right]$$

Interpretation (cont'd)

- For the case of truncation from below only, Greene (2000, p. 902) shows that

$$\frac{\partial E\left[y_i \mid y_i^* > a_L\right]}{\partial x} = \beta + \sigma \frac{\partial}{\partial x}\left[\lambda\left(\frac{\alpha_L}{\sigma}\right)\right]$$

$$= \beta\left[1 - \delta\left(\frac{\alpha_{Li}}{\sigma}\right)\right]$$

There is an *attenuation* of the effect of *x*

Example: Hausman and Wise (1977)

- Estimated earnings equation for low-income households (i.e., those below the poverty level)
- In this case, we have truncation from above
- Log(Earnings) were modeled as a function of
  - education
  - intelligence test scores
  - union membership
  - level of vocational training
  - disability status
  - age

Parameter Estimates

Variable	OLS	ML	Marg. Eff. (%)
Constant	8.203	9.102	NA
Education	0.010	0.015	0.36
IQ	0.002	0.006	0.14
Training	0.002	0.006	0.14
Union	0.090	0.246	NA
Illness	-0.076	-0.226	NA
Age	-0.003	-0.016	-0.39
σ	0.391	0.168	$\sigma_{y\|x} = 0.306$

Incidental Truncation

- Sometimes the truncations is based on variable other than the one being modeled
- Examples:
  - Recreation demand: on-site survey asking questions as to future visitation rates
  - Labor market: hours worked are observed only for individuals with a reservation wage exceeding market wage
  - Migration: income of immigrants, who are observed only if net benefit of moving is positive

Conditional Distributions

- We have two underlying variables

$$z_i^* = \gamma' w_i + u_i$$

$$y_i^* = \beta' x_i + \varepsilon_i$$

$$y_i = \begin{cases} y_i^* & z_i^* \in A \\ \text{no observations} & z_i^* \notin A \end{cases}$$

- Suppose, for simplicity, that $A = [a, \infty)$
- The truncated joint density of interest becomes

$$f\left(y_i^*, z_i^* \mid z_i^* \geq a\right) = \frac{f\left(y_i^*, z_i^*\right)}{\Pr\left(z_i^* \geq a\right)}$$

Moments of the Truncated Bivariate Normal Distribution

If

$$\begin{pmatrix} y \\ z \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_y \\ \mu_z \end{pmatrix}, \begin{pmatrix} \sigma_y^2 & \sigma_{yz} \\ \sigma_{yz} & \sigma_z^2 \end{pmatrix}\right)$$

$$E\left[y \mid z > a\right] = \mu_y + \rho\sigma_y\lambda\left(\alpha_z\right)$$

$$Var\left[y \mid z > a\right] = \sigma_y^2\left[1 - \rho^2\delta\left(\alpha_z\right)\right]$$

where

$$\alpha_z = \frac{\left(a - \mu_z\right)}{\sigma_z} \qquad \lambda\left(\alpha_z\right) = \frac{\phi\left(\alpha_z\right)}{1 - \Phi\left(\alpha_z\right)}$$

$$\delta\left(\alpha_z\right) = \lambda\left(\alpha_z\right)\left[\delta\left(\alpha_z\right) - \alpha_z\right] \qquad \rho = Corr\left(y, z\right)$$

Censoring

- Recall that, with censoring, the problem lies not with missing observations, but limitations on the available data
- Consider a latent variable of interest modeled as a linear function of observable characteristics

$$y_i^* = \beta' x_i + \varepsilon_i$$

- Observe

$$y_i = \begin{cases} y_i^* & y_i^* \in A \\ h\left(y_i^*\right) & y_i^* \notin A \end{cases}$$

Example: $A = [0, \infty)$ and $h\left(y_i^*\right) = 0 \forall y_i^* \notin A$

$$y_i = \begin{cases} y_i^* & y_i^* \geq 0 \\ 0 & y_i^* < 0 \end{cases}$$

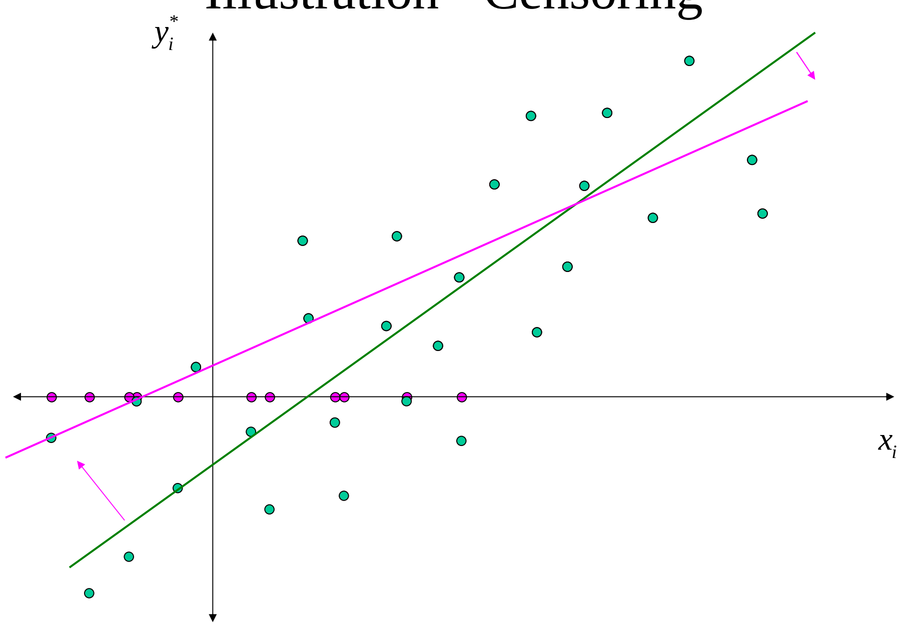
Examples (252 “hits” in EconLit)

- Capacity constrained data (e.g., ticket sales)
- Hours worked (or leisure demand) – essentially capacity constrained
- Commodity purchases (non-negative) (e.g., vacations)
- Number of arrests after release from prison
- Number of extramarital affairs
- Contributions to
  - Charity
  - Pension plans
- Lending practices
- Asset prices under price fluctuation limits
- R&D expenditures

Sources – Truncation Univariate Case

- *Greene, W. H., (2000) *Econometric Analysis*, 4th edition, Upper Saddle River, New Jersey: Prentice-Hall, Inc., Sections 20.3.
- Ruud, P., (2000) *An Introduction to Classical Econometric Theory*, New York: Oxford University Press, Ch. 28.
- Maddala, G. S., (1983) *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge: Cambridge University Press, Ch. 6.
- Mittelhammer, R. C., G. G. Judge, and D. J. Miller (2000), *Econometric Foundations*, Cambridge, MA: Cambridge University Press, Section 20.4.
- *Chib, S., (1992), "Bayes Inference in the Tobit Censored Regression Model," *Journal of Econometrics*, **51**: 79-99.
- *Amemiya, T., (1984), "Tobit Models: A Survey," *Journal of Econometrics* **24**: 3-61.

Illustration - Censoring



The Censored Distribution

- Analysis of censored data, as with truncated data, centers on understanding the distributional implications
- Consider the simple case of censoring from below; i.e.,

$$y_i^* = \beta' x_i + \varepsilon_i$$

$$y_i = \begin{cases} y_i^* & y_i^* \geq a_L \\ a_L & y_i^* < a_L \end{cases}$$

Tobit model corresponds to a_L=0

- Let

$$f_\varepsilon(\varepsilon_i) \text{ denote the pdf of } \varepsilon_i$$
$$F_\varepsilon(\varepsilon_i) \text{ denote the cdf of } \varepsilon_i$$

The Censored Distribution (cont'd)

- The pdf of y_i is a mixed probability function, with discrete and continuous components
- Ruud (2000, pp. 795-796) suggests deriving the pdf by first considering the corresponding cdf

For $c<a_L$: $\Pr[y_i < c \mid x_i] \leq \Pr[y_i < a_L \mid x_i] = 0$

For $c=a_L$: $\Pr[y_i = c \mid x_i] = \Pr[y_i^* < a_L \mid x_i] = \Pr[\varepsilon_i < a_L - \beta' x_i \mid x_i]$

$$= F_\varepsilon(a_L - \beta' x_i)$$

For $c>a_L$: $\Pr[y_i \leq c \mid x_i] = \Pr[y_i^* \leq c \mid x_i] = \Pr[\varepsilon_i \leq c - \beta' x_i \mid x_i]$

$$= F_\varepsilon(c - \beta' x_i)$$

The Censored Distribution (cont'd)

- The cdf of y_i is a then

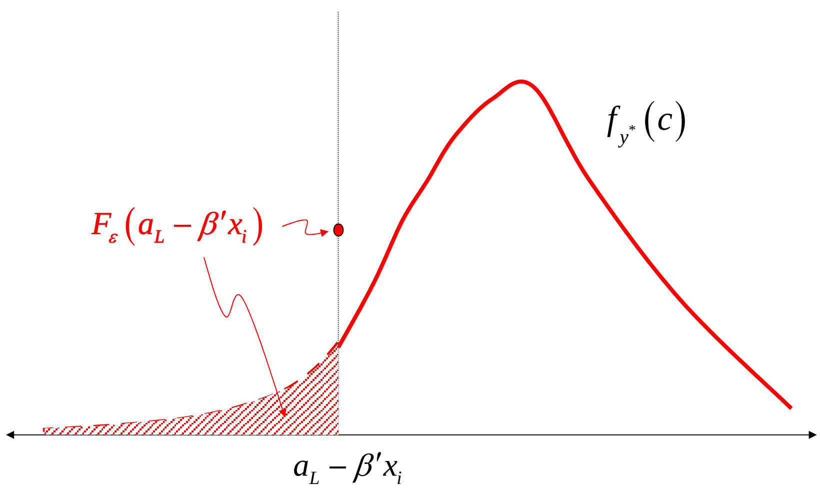
$$F_{y_i}(c \mid x_i) = \begin{cases} 0 & c < a_L \\ F_\varepsilon(c - \beta' x_i) & c \geq a_L \end{cases}$$

- The pdf is obtained by differentiating where F is differentiable and differencing where F jumps discretely

$$f_{y_i}(c \mid x_i) = \begin{cases} 0 & c < a_L \\ F_\varepsilon(a_L - \beta' x_i) & c = a_L \\ f_\varepsilon(c - \beta' x_i) & c > a_L \end{cases}$$

This pdf is used in specifying log-likelihood function for ML

Graphically



Moments of Censored Distribution

- In general

$$E[y_i | x_i] = E[y_i | y_i = a_L]\Pr[y_i = a_L] + E[y_i | y_i > a_L]\Pr[y_i > a_L]$$
$$= a_L F_\varepsilon(a_L - \beta' x_i) + \{\beta' x_i + E[\varepsilon_i | \varepsilon_i > a_L - \beta' x_i]\}[1 - F_\varepsilon(a_L - \beta' x_i)]$$
$$= \beta' x_i + \varphi(a_L - \beta' x_i)$$
$$= \beta' x_i + \varphi_i$$

where

$$\varphi(z) = zF_\varepsilon(z) + E[\varepsilon_i | \varepsilon_i > z][1 - F_\varepsilon(z)] > 0$$

and

$$Var(y_i | x_i) < Var(y_i^* | x_i)$$

Moments of Censored Normal Distribution

- If $\varepsilon_i \sim N(0, \sigma^2)$, then

$$E[y_i | x_i] = a_L \Phi_i + \{\beta' x_i + \sigma\lambda_i\}[1 - \Phi_i]$$
$$= \beta' x_i + (a_L - \beta' x_i)\Phi_i + \sigma\lambda_i[1 - \Phi_i]$$

$$Var(y_i | x_i) = \sigma^2[1 - \Phi_i]\left[(1 - \delta_i) + (\alpha_i - \lambda_i)^2 \Phi_i\right] < \sigma^2$$

where

$$\alpha_i = \frac{a_L - \beta' x_i}{\sigma} \qquad \Phi_i = \Phi\left(\frac{a_L - \beta' x_i}{\sigma}\right)$$

$$\lambda_i = \frac{\phi(\alpha_i)}{1 - \Phi(\alpha_i)} \qquad \delta_i = \lambda_i^2 - \lambda_i \alpha_i$$

Estimation

- OLS is biased, since

$$E\left[y_i \mid x_i\right] = \beta' x_i + \varphi\left(a_L - \beta' x_i\right)$$
$$= \beta' x_i + \varphi_i$$

$$\Rightarrow$$

$$E\left[b \mid x\right] = E\left[\left(x'x\right)^{-1} x'y \mid x\right]$$
$$= \left(x'x\right)^{-1} x'E\left[y \mid x\right]$$
$$= \beta + \left(x'x\right)^{-1} x'\varphi$$

As in the case of truncation, we will tend to get attenuation in the estimated parameters

Nonlinear Least Squares

- Assuming normal errors, NLS can be used, solving

$$\left(\hat{\beta}_{NLS}, \hat{\sigma}_{NLS}\right) = \arg\min_{\beta,\sigma}\left\{\left(y_i - E\left[y_i \mid x_i\right]\right)^2\right\}$$
$$= \arg\min_{\beta,\sigma}\left\{\left(y_i - \left[a_L \Phi_i + \left\{\beta' x_i + \sigma\lambda_i\right\}\left[1 - \Phi_i\right]\right]\right)^2\right\}$$

- This ignores heteroskedasticity
- Feasible Weight NLS (FWNLS) can be used to compensate

Heckman's (1976) Two-Step Procedure

- Heckman suggested a consistent estimator for the tobit model using
  - a probit model of censoring
  - LS applied using fitted components of conditional means
- To simplify the exposition, assume that

$$\varepsilon_i \sim N\left(0, \sigma^2\right)$$
$$a_L = 0$$

Heckman's (1976) Two-Step Procedure

- Recall that if $\varepsilon_i \sim N\left(0, \sigma^2\right)$, then censoring from below yields

$$E\left[y_i \mid x_i\right] = a_L \Phi\left(\frac{a_L - \beta' x_i}{\sigma}\right) + \left\{\beta' x_i + \sigma\lambda_i\right\}\left[1 - \Phi\left(\frac{a_L - \beta' x_i}{\sigma}\right)\right]$$

$$= \beta' x_i \Phi\left(\frac{\beta' x_i}{\sigma}\right) + \sigma\lambda_i \Phi\left(\frac{\beta' x_i}{\sigma}\right) \qquad \text{when } a_L = 0$$

$$= \beta' x_i \Phi\left(\frac{\beta' x_i}{\sigma}\right) + \sigma\phi\left(\frac{\beta' x_i}{\sigma}\right)$$

$$= \beta' x_i \Phi\left(\tilde{\beta}' x_i\right) + \sigma\phi\left(\tilde{\beta}' x_i\right)$$

$$= \beta' \tilde{x}_i + \sigma\phi_i$$

where $\tilde{x}_i = x_i \Phi\left(\tilde{\beta}' x_i\right)$ $\qquad \tilde{\beta} \equiv \sigma^{-1}\beta$

Heckman's (1976) Two-Step Procedure (cont'd)

- We then have a regression model

$$y_i = E(y_i \mid x_i) + \left[y_i - E(y_i \mid x_i) \right]$$
$$= \beta' \tilde{x}_i + \sigma\phi_i + \eta_i$$
$$= \alpha' w_i + \eta_i$$

where

$$\alpha = \begin{pmatrix} \beta \\ \sigma \end{pmatrix} \qquad w_i = \begin{pmatrix} \tilde{x}_i \\ \phi_i \end{pmatrix} \qquad \eta_i = y_i - E[y_i \mid x_i] \qquad E[\eta_i \mid x_i] = 0$$

- Given observations on w_i, estimates of β and σ can be obtained through OLS, but we don't have these

Heckman's (1976) Two-Step Procedure: Step 1

- The first step in Heckman's procedure estimates a probit model based on the discrete variable

$$\tilde{y}_i = \begin{cases} 1 & y_i^* \geq 0 \\ 0 & y_i^* < 0 \end{cases}$$

$$= \begin{cases} 1 & \varepsilon_i \geq -\beta' x_i \\ 0 & \varepsilon_i < -\beta' x_i \end{cases}$$

The corresponding log-likelihood function is

$$L(\tilde{y}, X, \tilde{\beta}) = \sum_i \left[(\tilde{y}_i) \ln\left(\Phi\left[\tilde{\beta}' x_i\right]\right) + (1 - \tilde{y}_i) \ln\left(1 - \Phi\left[\tilde{\beta}' x_i\right]\right) \right]$$

Heckman's (1976) Two-Step Procedure: Step 1 (cont'd)

- ML estimates from the first stage provide consistent estimates of $\tilde{\beta}$; i.e., the normalized parameters

- These can in turn be used to construct consistent estimators for $\tilde{x}_i$ and ϕ_i

$$\hat{\tilde{x}}_i = x_i \Phi\left(\hat{\tilde{\beta}} x_i\right)$$

$$\hat{\phi}_i = \phi\left(\hat{\tilde{\beta}} x_i\right)$$

Heckman's (1976) Two-Step Procedure: Step 2

- We can then write

$$\begin{aligned} y_i &= \beta' \tilde{x}_i + \sigma \phi_i + \eta_i \\ &= \beta' \hat{\tilde{x}}_i + \sigma \hat{\phi}_i + \beta'\left(\tilde{x}_i - \hat{\tilde{x}}_i\right) + \sigma\left(\phi_i - \hat{\phi}_i\right) + \eta_i \\ &= \alpha' \hat{w} + \hat{\eta}_i \end{aligned}$$

where

$$\hat{\eta}_i = \beta'\left(\tilde{x}_i - \hat{\tilde{x}}_i\right) + \sigma\left(\phi_i - \hat{\phi}_i\right) + \eta_i$$

- OLS will yield consistent, but not efficient, estimates of β and σ

- Conventional standard errors will be incorrect

Variation on Heckman's (1976) Two-Step Procedure: Step 2

- In the second stage, one can use just the uncensored observations
- Recall that

$$E\left[y_i \mid y_i > 0\right] = \beta' x_i + E\left[\varepsilon_i \mid \varepsilon_i > -\beta' x_i\right]$$
$$= \beta' x_i + \sigma\lambda_i$$

where

$$\lambda_i = \frac{\phi\left(\tilde{\beta}' x_i\right)}{\Phi\left(\tilde{\beta}' x_i\right)}$$

The second stage regression becomes

$$y_i = \beta' x_i + \sigma\hat{\lambda}_i + \hat{\eta}_i$$

ML for Censored Regression Model

- Recall that, when we have censoring from below:

$$f_{y_i}\left(c \mid x_i\right) = \begin{cases} 0 & c < a_L \\ F_\varepsilon\left(a_L - \beta' x_i\right) & c = a_L \\ f_\varepsilon\left(c - \beta' x_i\right) & c > a_L \end{cases}$$

- The corresponding log-likelihood function becomes:

$$LL\left(\beta \mid y, x\right) = \sum_{y_i = a_L} \ln\left[F_\varepsilon\left(a_L - \beta' x_i\right)\right] + \sum_{y_i > a_L} \ln\left[f_\varepsilon\left(y_i - \beta' x_i\right)\right]$$

ML for Censored Regression Model with Normal Errors

- When the errors are normal, the log-likelihood function becomes:

$$LL(\beta,\sigma \mid y,x) = \sum_{y_i=a_L} \ln\left[\Phi\left(\frac{a_L-\beta' x_i}{\sigma}\right)\right] + \sum_{y_i>a_L} \ln\left[\sigma^{-1}\phi\left(\frac{y_i-\beta' x_i}{\sigma}\right)\right]$$

$$= \sum_{y_i=a_L} \ln\left[\Phi\left(\frac{a_L-\beta' x_i}{\sigma}\right)\right] - \tfrac{1}{2}\sum_{y_i>a_L}\left\{\log(2\pi) + 2\log(\sigma) + \left(\frac{y_i-\beta' x_i}{\sigma}\right)^2\right\}$$

$$= \sum_{y_i=a_L} \ln\left[\Phi(\theta a_L - \gamma' x_i)\right] - \tfrac{1}{2}\sum_{y_i>a_L}\left\{\log(2\pi) - 2\log(\theta) + (\theta y_i - \gamma' x_i)^2\right\}$$

$$= LL(\theta,\gamma \mid y,x) \qquad \theta = \sigma^{-1} \text{ and } \gamma = \sigma^{-1}\beta$$

ML for Censored Regression Model with Normal Errors (cont'd)

- The reparameterized log-likelihood function is globally concave, insuring that local maximum is a global maximum (Olsen, 1978, *Econometrica*).
- The original parameters can then be obtained through the transformation

$$\sigma = \theta^{-1}$$

$$\beta = \frac{\gamma}{\theta}$$

with their asymptotic distributions characterized through simulation or Taylor-series approximations

Example: Female Labor Supply Quester and Greene (1978)

- Based on a random sample of 2798 wives questioned in 1967 U.S. Bureau of Census Survey
- Later used by Greene (1981), *Econometrica* to investigate asymptotic bias in OLS
  - Investigating empirical regularity that

$$\beta_{ML} \approx \frac{b}{P_1} \text{ where } P_1 = \text{proportion of uncensored observations}$$

Estimates

$P_1 = 0.46$

Variable	OLS	ML-Tobit	Corrected OLS
Constant	-690	-2754	-2,825
small children	-353	-824	-767
health	-412	-1010	-895
other income	0.73	1.03	1.59
wage	557	1,027	1211
south	275	588	598
farm	-204	-451	-444
urban	48.2	110	105
age	8.64	19.3	18.8
education	11.5	22.6	24.9
relative wage	124	286	269
2^{nd} marriage	13.1	25.3	28.6
mean divorce Pr.	219	481	477
high divorce Pr.	244	578	531

Alternative Estimation Procedures: EM Algorithm

- Recall, with the EM algorithm, an iterative process is used in which

E-Step: Form $H\left(\beta \middle| \hat{\beta}^t, y, x\right) = E\left[\sum_{i=1}^{N} \ln f\left(y_i^* \middle| \beta, y_i, x_i\right)\right].$

M-Step: Solve $\hat{\beta}^{t+1} = \arg\max_{\beta} H\left(\beta \middle| \hat{\beta}^t, y, x\right).$

- For the Tobit model:

E-step replaces censored observations with $\hat{y}_{i0} = E\left[y_i^* \mid y_i^* < 0, x_i\right]$

M-step corresponds to OLS with dependent variable $\left(y_1', \hat{y}_0'\right)'$

y_1 denotes uncensored observations, $\hat{y}_0$ denotes fitted conditional means

Alternative Estimation Procedures: Bayes Estimation

- Source: Chib, S., (1992), "Bayes Inference in the Tobit Censored Regression Model," *Journal of Econometrics*, **51**: 79-99.
- Relies on data augmentation combined with Gibbs sampling – similar to EM algorithm

Single Site Model of Recreation Demand

Suppose we have the following latent variable regression model:

$$y_i^* = x_i'\beta + \varepsilon_i; \quad \varepsilon_i \sim N\left(0, \tau^{-2}\right)$$

and observed variable is:

$$y_i = \begin{cases} y_i^* & y_i^* \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

The corresponding likelihood function is given by:

$$LL\left(y \middle| \beta, \tau\right) = \sum_{y_i=0} \ln\left[1 - \Phi\left(x_i'\beta\tau\right)\right]$$

$$-\frac{n_1}{2}\left[\log\left(2\pi\right) - 2\log\left(\tau\right)\right] - \tfrac{1}{2}\sum_{y_i > a_L} \tau^2\left(y_i - x_i'\beta\right)^2$$

Bayesian Analysis of the Tobit Model Using Data Augmentation

- Bayesian analysis would then combine data with prior information to form posterior

$$p\left(\theta \middle| y, X\right) \propto p\left(\theta\right) p\left(y \middle| X, \theta\right)$$

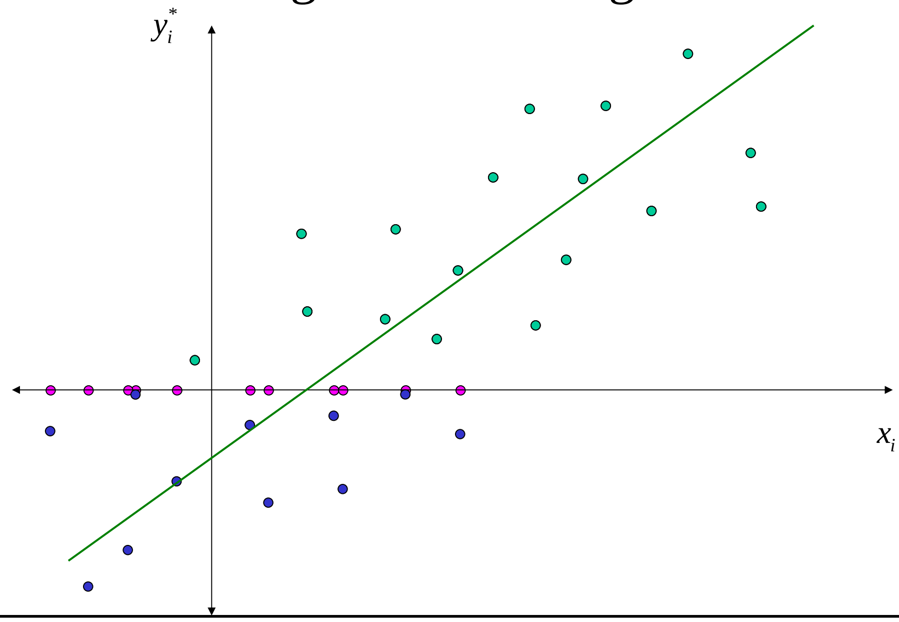
difficult to directly characterize

- Let z denote those elements of y^* that are censored and

$$y_1 = \left\{y_i \middle| y_i^* > 0\right\} \qquad y_0 = \left\{y_i \middle| y_i^* \leq 0\right\}$$

- Chib (1992) suggests converting the Tobit model to a simple linear model by simulating z

Basic Insight of Data Augmentation



Generic Gibbs Sampling

Given a generic multivariate density function of interest, say

$$p\left(\theta_1,\theta_2,\ldots,\theta_K\right)$$

And known conditional density functions

$$p\left(\theta_k \middle| \theta_j \forall j \neq k\right)$$

The sequence

$$\left\{\left(\theta_1^t,\theta_2^t,\ldots,\theta_K^t\right) \text{ where } \theta_k^t \sim p\left(\theta_k \middle| \theta_1^t,\ldots,\theta_{k-1}^t,\theta_{k+1}^{t-1},\ldots,\theta_K^{t-1}\right)\right\}$$

$$\xrightarrow{d} p\left(\theta_1,\theta_2,\ldots,\theta_K\right)$$

Gibbs Sampling Routine for the Tobit model

We want to characterize $p\left(\beta,\tau,z \mid y_1,x\right)$

The trick in Chib's method is to use Gibb's sampling, sequentially drawing from

$$p\left(z \mid \beta,\tau,y_1,x\right)$$

$$p\left(\beta \mid z,\tau,y_1,x\right) = p\left(\beta \mid y^*,\tau,x\right)$$

$$p\left(\tau \mid z,\beta,y_1,x\right) = p\left(\tau \mid y^*,\beta,x\right)$$

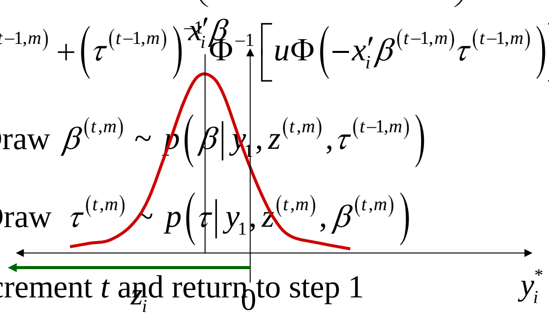
Gibbs Sampling Routine for the Tobit model

For $m = 1,\ldots,M; t = 0,\ldots,T$

Start with t=1 and starting values $\left(\beta^{(0,m)},\tau^{(0,m)}\right)$

- Step 1: Draw $z_i^{(t,m)} \sim TN\left(x_i'\beta^{(t-1,m)},\left(\tau^{(t-1,m)}\right)^{-2}\right)$

$$z_i^{(t,m)} = x_i'\beta^{(t-1,m)} + \left(\tau^{(t-1,m)}\right)^{-1}\Phi^{-1}\left[u\Phi\left(-x_i'\beta^{(t-1,m)}\tau^{(t-1,m)}\right)\right]; u \sim U\left[0,1\right]$$



- Step 2: Draw $\beta^{(t,m)} \sim p\left(\beta \mid y_1, z^{(t,m)},\tau^{(t-1,m)}\right)$

- Step 3: Draw $\tau^{(t,m)} \sim p\left(\tau \mid y_1, z^{(t,m)},\beta^{(t,m)}\right)$

If $t<T$, increment t and return to step 1

Gibbs Sampling Routine (cont'd)

- The M cycles produce M draws $\left(\beta^{(T,m)}, \tau^{(T,m)}\right); m = 1, \ldots, M$, from the posterior distribution of interest.
- Marginal posterior densities can be constructed as:

$$\hat{p}\left(\beta \middle| y\right) = \tfrac{1}{M} \sum_{m=1}^{M} p\left(\beta \middle| y, z^{(T,m)}, \tau^{(T,m)}\right)$$

$$\hat{p}\left(\tau \middle| y\right) = \tfrac{1}{M} \sum_{m=1}^{M} p\left(\tau \middle| y, z^{(T,m)}, \beta^{(T,m)}\right)$$

Gibbs Sampling Routine with Diffuse Prior

Let: $p(\beta, \tau) \propto \tau^{-2}$

$$p\left(z_i \middle| \beta, \tau, y_i = 0\right) = f_{TN}\left(x_i'\beta, \tau^{-2}\right)$$

$$p\left(\beta \middle| y_1, z, \tau\right) = f_N\left(\beta \middle| \hat{\beta}_z, \tau^{-2}\left(X'X\right)^{-1}\right)$$

$$\hat{\beta}_z = \left(X'X\right)^{-1} X'y_z$$

$$y_z = \begin{pmatrix} y_1 \\ z \end{pmatrix}$$

$$p\left(\tau^2 \middle| y_1, z, \beta\right) = f_G\left(\tau^2 \middle| \frac{n}{2}, \frac{e_z'e_z}{2}\right)$$

← Gamma distribution

$$e_z = y_z - X'\beta$$

Adding Prior Information

Suppose you have the following prior information

$$\beta \sim N\left(\beta_0, H^{-1}\right)$$

where *H* is a fixed precision matrix.

$$p\left(\beta \middle| y_1, z, \tau\right) = f_N\left(\beta \middle| \overline{\beta}_z, \left(\tau^2 X'X\right)^{-1}\right)$$

$$\overline{\beta}_z = \left(H + \tau^2 X'X\right)^{-1}\left(H\beta_0 + \tau^2 X'y_z\right) = W_P \beta_0 + W_D \beta_z$$

$$W_P \equiv \left(H + \tau^2 X'X\right)^{-1} H$$

$$W_S \equiv \left(H + \tau^2 X'X\right)^{-1} \tau^2 X'X$$

Marginal Effects

- In general, for a linear regression model censored from below:

$$\frac{\partial E\left(y_i^*\right)}{\partial x_i} = \beta$$

$$\frac{\partial E\left(y_i\right)}{\partial x_i} = \beta\left[1 - F_\varepsilon\left(\frac{a_L - \beta' x_i}{\sigma}\right)\right], \sigma^2 \equiv Var\left(\varepsilon_i \mid x_i\right)$$

which you are interested in depends upon application

Marginal Effects
Normal Case

- Given normal errors

$$\frac{\partial E(y_i)}{\partial x_i} = \beta[1-\Phi_i]$$
$$= \beta[1-\Phi_i]\left[1-\lambda_i(\alpha_i+\lambda_i)+\lambda_i(\alpha_i+\lambda_i)\right]$$
$$= [1-\Phi_i]\beta\left[1-\lambda_i(\alpha_i+\lambda_i)\right]+\beta[1-\Phi_i]\left[\lambda_i(\alpha_i+\lambda_i)\right]$$
$$= [1-\Phi_i]\beta\left[1-\lambda_i(\alpha_i+\lambda_i)\right]+\beta(\alpha_i+\lambda_i)\phi_i$$
$$= \Pr[y_i > a_L]\underbrace{\frac{\partial E[y_i \mid y_i > a_L]}{\partial x_i}}_{\text{change in conditional mean}} + E[y_i \mid y_i > a_L]\underbrace{\frac{\partial \Pr[y_i > a_L]}{\partial x_i}}_{\text{change in censoring prob.}}$$

Other Specification Issues

- Heteroskedasticity
- Misspecification of the censoring probability
- Nonnormality
- Conditional Moment Tests

Heteroskedasticity

- One approach to dealing with heteroskedasticity is to attempt to explicitly model it; i.e., let

$$\sigma_i^2 = \sigma^2 \exp\left(\alpha' w_i\right)$$

- Homoskedasticity corresponds to the restriction:

$$\alpha = 0$$

- One does not have to actually estimate heteroskedastic model, but can use Lagrange multiplier test
  See Greene, p. 914.
- One does, however, need to specify w_i's

Mispecification of Censoring Probability

- The tobit model assumes that the parameters determining the censoring are the same that determine the level of the latent variable without censoring.
- Counter example: Fin and Schmidt (1984) analyzing loss due to fires:
  - probability of fire is likely to increase with age
  - loss given a fire is likely to diminish with age

  In essence, they are suggesting an incidental censoring problem

Generalizations of Tobit

- Greene (2000) suggests
  1. Decision equation: $\Pr\left[y_i^* > 0\right] = \Phi(\gamma' x) \qquad z_i = 1$

     $\Pr\left[y_i^* > 0\right] = 1 - \Phi(\gamma' x) \qquad z_i = 0$
  2. Regression equation for nonlimit observations:

$$E\left[y_i \mid z_i = 1\right] = \beta' x_i + \rho \sigma_y \lambda_i(\alpha_z)$$

Generalizations of Tobit

- This is a form of incidental censoring

$$z_i^* = \gamma' w_i + u_i$$

$$y_i^* = \beta' x_i + \varepsilon_i$$

$$y_i = \begin{cases} y_i^* & z_i^* \geq 0 \\ 0 & z_i^* < 0 \end{cases}$$

$$E\left[y \mid z > a\right] = \beta' x_i + \rho \sigma_y \lambda(\alpha_z)$$

- Estimation can be done via ML or 2-step Heckman type procedure

Consumer Demand Systems with Non-Negativity Constraints

- There are numerous empirical settings in which non-negativity constraints are binding in models of consumer demand
  - Recreation demand
  - Food demand
  - Grant funding, etc.
- Four Fundamental Approaches
  - Discrete:
    - Repeated Multivariate Discrete Choice
    - Count Data Models
  - Continuous
    - Statistical: Amemiya-Tobin
      - Maximum Likelihood
      - Two-step procedures
    - Economic: Kuhn-Tucker
      - Primal
      - Dual

Repeated Multivariate Discrete Choice

- We have already seen one approach to non-negativity constraints in demand analysis
  - Model choices as a sequence of discrete decisions (or choice occasions)
  - Decisions can be correlated over choice occasions (e.g., through alternative mixing distributions in mixed logit)
- Limitations
  - Definition of "choice occasions" is artificial and likely endogenous
  - Traditionally, latent variables have been specified as simple linear reduced form equations (i.e., flexible functional forms rarely used)

Sources

- *Wales, T., and A. Woodland (1983), "Estimation of Consumer Demand Systems with Binding Non-Negativity Constraints," *Journal of Econometrics* **21**: 263-285.
- Amemiya, T., (1974), "Multivariate Regression and Simultaneous Equation Models when the Dependent Variables are Truncated Normal," *Econometrica* **42**: 999-1012.
- Lee, L., and M. Pitt (1986), "Microeconometric Demand Systems with Binding Non-negativity Constraints: The Dual Approach," *Econometrica* **54**: 1237-1242.

Sources (cont'd)

- *Phaneuf, D., C. Kling, and J. Herriges (2000), "Estimation and Welfare Calculations in a Generalized Corner Solution Model with an Application to Recreation Demand," *The Review of Economics and Statistics*, **82**: 83-92.
- *Phaneuf, D., (1999), "A Dual Approach to Modeling Corner Solutions in Recreation Demand," *Journal of Environmental Economics and Management*, **37**: 85-105.
- Heien, D., and C. Wessells (1990), "Demand Systems Estimation with Microdata: A Censored Regression Approach," *Journal of Business and Economic Statistics,* **8**: 365-71.
- *Shonkwiler, J. S., and S. Yen (1999), "Two-Step Estimation of a Censored Demand System of Equations," *American Journal of Agricultural Economics* **81**: 972-982.

The Amemiya-Tobin Approach

- The Amemiya-Tobin approach to non-negativity constraints in a system of demand equations is essentially an extension of Tobit to multiple dimensions
- The derivations in Wales and Woodland (1983) focus on system of share equations.
- Suppose we specify a deterministic indirect utility function associated with the consumption of M commodities

$$V\left(p_{i\bullet}, I_i\right)$$

where $p_{i\bullet}$ denotes the vector of commodity prices and I_i denotes income

The Amemiya-Tobin Approach (cont'd)

- Apply Roy's Identity to derive share equations:

$$s_{im} = \frac{p_i y_i}{I} = s_m\left(p_{i\bullet}, I_i\right) \; m = 1, \ldots, M$$

where

$$s_{im} \in [0,1] \text{ and } \sum_{m=1}^{M} s_{im} = 1$$

- The above deterministic share equations, however, do not have any random components.

The Amemiya-Tobin Approach (cont'd)

- Common practice is to convert the deterministic system of shares into a random system of shares by adding a random component to each share
- Let

$$y_{im}^{*} = s_i\left(p_{i\bullet}, I_i\right) + \varepsilon_{im};\, m = 1,\ldots, M-1$$

$$y_{iM}^{*} = 1 - \sum_{m=1}^{M-1} y_{im}^{*}$$

$$\varepsilon_{i,-M} = \left(\varepsilon_{i1},\ldots,\varepsilon_{iM-1}\right)' \sim N\left(0,\Omega\right)$$

The Amemiya-Tobin Approach (cont'd)

- The problem is that these share equations, while they sum to unit by construction, need not lie in the unit interval.
- One solution is to view the observed data as a mapping (specifically, a censoring) of latent variables such that

$$y_{im} = \begin{cases} \dfrac{y_{im}^{*}}{\sum_{n\in J_i} y_{in}^{*}} & y_{im}^{*} > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$J_i \equiv \left\{m : y_{im}^{*} > 0, m = 1,\ldots, M\right\}$$

i.e., "out of Simplex" censoring

The Amemiya-Tobin Approach (cont'd)

- Note that

$$y_{im} = \begin{cases} \dfrac{s_m\left(p_{i\bullet}, I_i\right) + \varepsilon_{im}}{\sum_{n \in J_i}\left[s_n\left(p_{i\bullet}, I_i\right) + \varepsilon_{in}\right]} & y_{im}^* > 0 \\ 0 & \text{otherwise} \end{cases}$$

so that all prices enter the share equations and the basic functional form does not depend on which elements are in J_i.

The Density Function for A-T

- Without loss of generality, suppose the first K elements of our observed demands are strictly positive and the rest are negative; i.e.,

$$y_{i\bullet} = \left(\frac{y_{i1}^*}{\sum_{m=1}^{K} y_{im}^*}, \ldots, \frac{y_{i1}^*}{\sum_{m=1}^{K} y_{im}^*}, 0, \ldots, 0\right)'$$

$$f\left(y_{i\bullet}\right) = \begin{cases} \displaystyle\int_{y_1}^{\infty}\int_{\alpha_{K+1}}^{0}\cdots\int_{\alpha_{M-1}}^{0} n\left[\sigma_1\left(y_1^*\right),\ldots,\sigma_K\left(y_K^*\right), y_{K+1}^*,\ldots,y_M^*\right] J\left(y\right) dy_{M-1}^*\cdots dy_{K+1}^* dy_1^* & 1 \le K < M \\ n\left(y_{i1}^*,\ldots,y_{iM-1}^*\right) & K = M \end{cases}$$

The Density Function for A-T (cont'd)

where

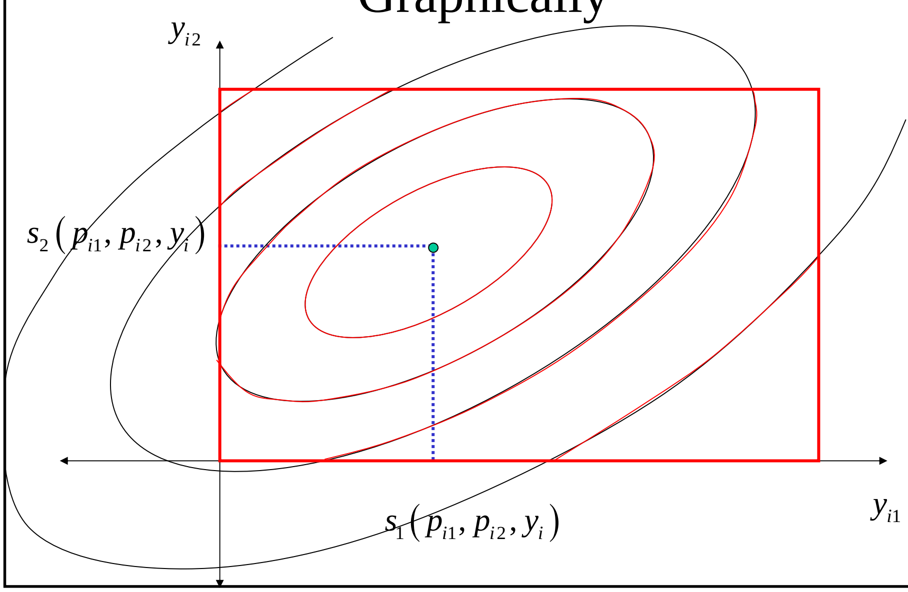
$$\alpha_{K+1} \equiv 1 - \sum_{k=1}^{K} \sigma_k \left(y_1^* \right)$$

$$\alpha_j \equiv \alpha_{K+1} - \sum_{k=K+1}^{j-1} y_k^* \qquad j = K+2, \dots, M-1$$

$$\sigma_j \left(y_{i1}^* \right) \equiv y_{i1}^* \left(y_{ij} / y_{i1} \right) \qquad j = 1, \dots, K$$

$$J(y) \equiv \left[\sum_{k=1}^{K} \left(\frac{y_{ik}}{y_{i1}} \right)^2 \right]^{\frac{1}{2}}$$

Graphically



Log-likelihood Function

- The corresponding log-likelihood function is given by

$$LL = \sum_{i=1}^{N} \ln f\left(y_{i\cdot}\right)$$

- Difficulties
  - The likelihood function is complex
    - involves multidimensional integrals
    - density function varies by individual, with as many as *M*!/[*K*!(*M*-*K*)!] variants for a given *K*
  - There is no economic interpretation as to the
    - source of the error
    - why it is censored at zero

Two-Step Procedures - HW

- Heien and Wessells (1990) suggest a two-step procedure analogous to Heckman's two-step single equation method

Step 1: For each equation, estimate a probit model of censoring; i.e., model

$$z_{im} = \begin{cases} 1 & y_{im}^* > 0 \\ 0 & y_{im}^* \le 0 \end{cases}$$

If, for example,

$$y_{im}^* = \beta_m' x_{im} + \varepsilon_{im}$$

Step 1 then yields estimates of $\tilde{\beta}_m = \sigma_m^{-1}\beta_m,\ \sigma_m^2 = Var\left(\varepsilon_{im}\right)$

Two-Step Procedures – HW (cont'd)

Step 2: Estimate SUR model using all of the data and the system of equations:

$$y_{im} = \beta_m' x_{im} + \sigma_m \hat{\lambda}_{im} + \varepsilon_{im}$$

where

$$\hat{\lambda}_i = \frac{\phi\left(d_{im}\hat{\tilde{\beta}}_m' x_{im}\right)}{\Phi\left(d_{im}\hat{\tilde{\beta}}_m' x_{im}\right)}$$

and

$$d_{im} = 2z_{im} - 1$$

Two-Step Procedures – HW (cont'd)

There are numerous problems with this procedure

1. It is incorrect, even for a single equation

Recall that for a single equation:

$$E\left[y_i \mid x_i\right] = \beta' x_i \Phi\left(\tilde{\beta}' x_i\right) + \sigma\phi\left(\tilde{\beta}' x_i\right)$$

Only if one uses only the uncensored data is it true that

$$E\left[y_i \mid y_i > 0\right] = \beta' x_i + E\left[\varepsilon_i \mid \varepsilon_i > -\beta' x_i\right]$$
$$= \beta' x_i + \sigma\lambda_i$$

The analogy applies for multiple equations

Two-Step Procedures – Shonkwiler and Yen (1999) fix

Shonkwiler and Yen (1999) (SY) propose a fix to HW

They propose proceeding with step 1, followed by a modified step 2, estimating the SUR system

$$y_{im} = \beta_m' \hat{\tilde{x}}_{im} + \sigma_m \hat{\phi}_{im}$$

where

$$\hat{\tilde{x}}_{im} = x_{im} \Phi\left(\hat{\tilde{\beta}}_m' x_{im}\right)$$

and

$$\hat{\phi}_{im} = \phi\left(\hat{\tilde{\beta}}_m' x_{im}\right)$$

SY-Procedure

SY note that the residual for this regression is given by

$$\xi_{im} = \varepsilon_{im} + \beta_m' \left(\tilde{x}_{im} - \hat{\tilde{x}}_{im}\right) + \sigma_m \left(\phi_{im} - \hat{\phi}_{im}\right)$$

Thus, the residuals are both heteroskedastic and correlated across equations and individuals

Conceptually, one could correct for both these factors, though it is not trivial to do so.

Other Problems with Two-Step Methods

- Neither HW nor SY is efficient, since they each ignore:
  - correlation across equations to begin with
    - They inefficiently estimate censoring probability by not estimating a system of probits
    - Efficiency is less of a problem if share equations are uncorrelated, but this is never the case since they sum to one
  - cross-equation constraints (e.g., if a parameter appears in more than one equation, it is separately estimated in each probit equation)

The Kuhn-Tucker Approach

- The Kuhn-Tucker model, developed originally by Wales and Woodland (1983) relies on an integrated economic model of both corners and interior solutions
- Begins with assumption that consumer have preferences over M+1 commodities, one of which is a numeraire good

The Kuhn-Tucker Approach

Consumer solves

$$\underset{x,z}{Max}\, U\left(x, z, q, \gamma, \varepsilon\right)$$
$$s.t. \quad p'x + z \le y$$
$$z \ge 0$$
$$x \ge 0$$

where

$x = \left(x_1, \ldots, x_M\right)'$ denotes the vector of goods

$p = \left(p_1, \ldots, p_M\right)'$ denotes the vector of prices

$q = \left(q_1, \ldots, q_M\right)'$ denotes the vector of good attributes

$\varepsilon = \left(\varepsilon_1, \ldots, \varepsilon_M\right)'$ is a vector of random components capturing preference variation

y and *z* denote income and the numeraire good, respectively

The Kuhn-Tucker Approach (cont'd)

- The utility function is assumed to be a quasi-concave, increasing, and continuously differentiable function of (*x*,*z*).
- The corresponding first-order necessary and sufficient Kuhn-Tucker conditions are

$$U_j\left(x, z; q, \gamma, \varepsilon\right) \equiv \frac{\partial U\left(x, z; q, \gamma, \varepsilon\right)}{\partial x_j} \le \lambda p_j, \quad x_j \ge 0,$$
$$x_j\left[U_j\left(x, z; q, \gamma, \varepsilon\right) - \lambda p_j\right] = 0, \quad j = 1, \ldots, M$$
$$U_z\left(x, z; q, \gamma, \varepsilon\right) \equiv \frac{\partial U\left(x, z; q, \gamma, \varepsilon\right)}{\partial z} \le \lambda, \quad z \ge 0$$
$$z\left[U_z\left(x, z; q, \gamma, \varepsilon\right) - \lambda\right] = 0, \quad j = 1, \ldots, M$$

The Kuhn-Tucker Approach (cont'd)

and

$$p'x + z \le y, \qquad \lambda \ge 0, \qquad \lambda\left(y - p'x - z\right)$$

with λ denoting the marginal utility of income.

If we assume that the numeraire good is a necessary good, then:

$$\lambda = U_z\left(x, z; q, \gamma, \varepsilon\right)$$

and

$$z = y - p'x$$

The Kuhn-Tucker Conditions

Substituting, the KT conditions become

$$U_j\left(x, y - p'x; q, \gamma, \varepsilon\right) \le U_z\left(x, y - p'x; q, \gamma, \varepsilon\right) p_j$$

$$x_j \ge 0,$$

$$x_j\left[U_j\left(x, y - p'x; q, \gamma, \varepsilon\right) - U_z\left(x, y - p'x; q, \gamma, \varepsilon\right) p_j\right] = 0$$

KT Assumptions on Errors

Assume that

$$U_{z\varepsilon} = 0$$

$$\frac{\partial U_j}{\partial \varepsilon_k} = 0 \quad \forall k \neq j$$

$$\frac{\partial U_j}{\partial \varepsilon_j} > 0 \quad \forall j$$

$\Rightarrow$

$$U_j\left(x, y - p'x; q, \gamma, \varepsilon\right) = \tilde{U}_j\left(x, y - p'x; q, \gamma, \varepsilon_j\right)$$

$$U_z\left(x, y - p'x; q, \gamma, \varepsilon\right) = \tilde{U}_z\left(x, y - p'x; q, \gamma\right)$$

Implicit Solutions

Let

$$g_j = g_j\left(x, y, p; q, \gamma\right)$$

implicitly solve

$$\tilde{U}_j\left(x, y - p'x; q, \gamma, g_j\right) - \tilde{U}_z\left(x, y - p'x; q, \gamma\right) p_j = 0$$

Our first order conditions become

$$\varepsilon_j \leq g_j\left(x, y - p'x; q, \gamma\right)$$

$$x_j \geq 0,$$

$$x_j\left[\varepsilon_j - g_j\left(x, y - p'x; q, \gamma\right)\right] = 0$$

This provides the information needed to define log-likelihood

Contribution to Log-Likelihood

Consider an individual who chooses to consume positive quantities for first K commodities

$$x_k > 0 \Rightarrow \varepsilon_k = g_k\left(x, y, p; q, \gamma\right) \qquad k = 1, \ldots, K$$

$$x_k = 0 \Rightarrow \varepsilon_k \leq g_k\left(x, y, p; q, \gamma\right) \qquad k = K+1, \ldots, M$$

This individual's contribution to the likelihood function is

$$\int_{-\infty}^{g_{K+1}} \int_{-\infty}^{g_{K+2}} \cdots \int_{-\infty}^{g_M} f_\varepsilon\left(g_1, \ldots, g_K, \varepsilon_{K+1}, \ldots, \varepsilon_M\right) abs\left|J_K\right| d\varepsilon_{K+1} d\varepsilon_{K+2} \cdots d\varepsilon_M$$

J_k denotes Jacobian for transformation from ε to $\left(x_1, \ldots, x_K, \varepsilon_{K+1}, \ldots, \varepsilon_M\right)'$

Utility Maximization as Two-Stage Optimization

Let $A = \left\{\varnothing, \{1\}, \{2\}, \ldots, \{M\}, \{1,2\}, \{1,3\}, \ldots, \{1, 2, \ldots, M\}\right\}$ denote the collection of all possible subsets of the index set $I=\{1,\ldots,M\}$.

Construct a conditional utility function for each $\omega \in A$

$$V_\omega\left(p_\omega, y, q, \gamma, \varepsilon\right) \equiv \underset{x,z}{Max}\, U\left(x, z, q, \gamma, \varepsilon\right)$$

$$\begin{aligned} s.t. \quad & \sum_{j \in \omega} p_j x_j + z \leq y \\ & z \geq 0 \\ & x_j = 0 \quad j \notin \omega \\ & x_j \geq 0 \quad j \in \omega \end{aligned}$$

$$p_\omega = \left\{p_j : j \in \omega\right\}$$

Utility Maximization as Two-Stage Optimization

Note: $V_\omega\left(p_\omega, y, q, \gamma, \varepsilon\right)$ does not depend on prices of unconsumed goods

Let

$x_{\omega j}\left(p_\omega, y, q, \gamma, \varepsilon\right)$ denote conditional demand levels

$$\tilde{A} = \tilde{A}\left(p, y; q, \gamma, \varepsilon\right)$$
$$= \left\{\omega \in A \mid x_{\omega j}\left(p_\omega, y; q, \gamma, \varepsilon\right) > 0 \ \forall j \in \omega\right\}$$

Then

$$V\left(p, y; q, \gamma, \varepsilon\right) = \underset{\omega \in A}{Max}\left\{V_\omega\left(p_\omega, y; q, \gamma, \varepsilon\right)\right\}$$
$$= \underset{\omega \in \tilde{A}}{Max}\left\{V_\omega\left(p_\omega, y; q, \gamma, \varepsilon\right)\right\}$$

Application – Phaneuf, Kling and Herriges (2000)

- Recreation demand for anglers in the Wisconsin Great Lakes region.
- Based on survey of fishing license holders
- 22 fishing destinations were aggregated into four sites:
  - Site 1: Lake Superior
  - Site 2: South Lake Michigan
  - Site 3: North Lake Michigan
  - Site 4: Green Bay
- Prices for aggregate sites were constructed as weighted averages of disaggregate round-trip costs
- Catch rate data were available by site and species
- Toxin Levels were also available by aggregate sites

Empirical Specification

- The direct utility function was assumed to be a variant of the linear expenditure system

$$U\left(x,z;q,\gamma,\varepsilon\right)=\sum_{j=1}^{M}\Psi_j\left(q_j,\varepsilon_j\right)\ln\left(x_j+\Omega\right)+\ln\left(z\right)$$

where

$$\Psi_j\left(q_j,\varepsilon_j\right)-\exp\left(\sum_{k=1}^{K}\delta_k q_{jk}+\varepsilon_j\right)\qquad j=1,\ldots,M$$

The Ψ_j's can be thought of as quality indices

$\gamma=\left(\delta,\Omega\right)$ denotes the parameters to be estimated

Empirical Specification (cont'd)

- Recall that our first order KT conditions can be written as

$$\varepsilon_j\le g_j\left(x,y-p'x;q,\gamma\right)\quad x_j\ge 0,$$

$$x_j\left[\varepsilon_j-g_j\left(x,y-p'x;q,\gamma\right)\right]=0$$

where $g_j=g_j\left(x,y,p;q,\gamma\right)$ implicitly solves

$$\tilde{U}_j\left(x,y-p'x;q,\gamma,g_j\right)-\tilde{U}_z\left(x,y-p'x;q,\gamma\right)p_j=0$$

For our empirical specification, one can explicitly derive

$$g_j\left(x,y,p;q,\gamma\right)=\ln\left[\frac{p_j\left(x_j+\Omega\right)}{y-\sum_{j=1}^{M}p_jx_j}\right]-\sum_{k=1}^{K}\delta_k q_{jk}$$

Error Specification

- The errors were assumed to be drawn from a GEV distribution, with nesting structure {(1,2),(3,4)}
- The paper also presents results using an EV error structure
- Subsequent papers have employed a error specifications analogous to mixed logit terms used in Phaneuf and Herriges (2001).

Log-Likelihood

- The resulting log-likelihood function has the form

$$
\begin{aligned}
\ln \mathrm{L} = & -\sum_{j=1}^{2} \frac{d_j g_j}{\mu\theta^S} - \sum_{j=3}^{4} \frac{d_j g_j}{\mu\theta^N} \\
& - \left[\exp\left(\frac{-g_1}{\mu\theta^S}\right) + \exp\left(\frac{-g_2}{\mu\theta^S}\right) \right]^{\theta^S} - \left[\exp\left(\frac{-g_3}{\mu\theta^N}\right) + \exp\left(\frac{-g_4}{\mu\theta^N}\right) \right]^{\theta^N} \\
& + (d_1 + d_2)(\theta^S - 1) \log\left[\exp\left(\frac{-g_1}{\mu\theta^S}\right) + \exp\left(\frac{-g_2}{\mu\theta^S}\right) \right] \\
& + (d_3 + d_4)(\theta^N - 1) \log\left[\exp\left(\frac{-g_3}{\mu\theta^N}\right) + \exp\left(\frac{-g_4}{\mu\theta^N}\right) \right] \\
& + (d_1 d_2) \log\left(1 - \frac{\theta^S - 1}{\theta^S} \left[\exp\left(\frac{-g_1}{\mu\theta^S}\right) + \exp\left(\frac{-g_2}{\mu\theta^S}\right) \right]^{-\theta^S} \right) \\
& + (d_3 d_4) \log\left(1 - \frac{\theta^N - 1}{\theta^N} \left[\exp\left(\frac{-g_3}{\mu\theta^N}\right) + \exp\left(\frac{-g_4}{\mu\theta^N}\right) \right]^{-\theta^N} \right) \\
& + \log\left| J_\omega \right| + (d_1 + d_2 + d_3 + d_4) \log(\mu),
\end{aligned}
$$

Log-Likelihood (cont'd)

where

$$d_i = \begin{cases} 1 & x_i > 0 \\ 0 & otherwise \end{cases}$$

and J_ω denotes the Jacobian transformations, which vary by consumption configurations (i.e., ω)

$$J_\omega = F_j \text{ for } \omega = \{j\},\ j = 1,2,3,4.$$

$$F_j \equiv \frac{1}{x_j + \theta} + z_j$$

$$z_j \equiv \frac{p_j}{y - \sum_{k=1}^{4} p_k x_k}$$

Log-Likelihood (cont'd)

$$J_\omega = \prod_{j\in\omega} F_j - \prod_{j\in\omega} z_j \text{ for } \omega = \{1,2\},\{1,3\},\{1,4\},\{2,3\},\{2,4\}, \text{ and } \{3,4\}.$$

$$J_\omega = \prod_{j\in\omega} F_j + 2\prod_{j\in\omega} z_j - \sum_{j\in\omega} F_j \left(\prod_{\substack{k\in\omega \\ k\neq j}} z_k \right)$$

$$\text{for } \omega = \{1,2,3\},\{1,2,4\},\{1,3,4\}, \text{ and } \{2,3,4\}.$$

$$J_\omega = F_1F_2F_3F_4 - 3z_1z_2z_3z_4 + 2\left(F_1z_2z_3z_4 + F_2z_1z_3z_4 + F_3z_1z_2z_4 + F_4z_1z_2z_3\right)$$
$$\text{-}\left(F_1F_2z_3z_4 + F_1F_3z_2z_4 + F_1F_4z_2z_3 + F_2F_3z_1z_4 + F_2F_4z_1z_3 + F_3F_4z_1z_2\right)$$
$$\text{for } \omega = \{1,2,3,4\}$$

Parameter Estimates

Parameter	EV errors		GEV Errors	
	Estimate	P-value	Estimate	P-value
δ_0 (Intercept)	-8.53	<0.001	-8.43	<0.001
δ_{lk} (Lake Trout)	0.10	0.953	-0.70	0.667
δ_{ch} (Chinook S.)	13.39	<0.001	11.11	<0.001
δ_{co} (Coho S.)	3.12	0.023	3.71	0.007
δ_{rb} (Rainbow T.)	8.61	0.035	13.96	<0.001
δ_E (Eff. Toxin)	-0.06	0.018	-0.07	<0.001
Ω	1.76	<0.001	1.82	<0.001
θ^N sites (3,4)	1.00	NA	0.57	<0.001
θ^S sites (1,2)	1.00	NA	0.92	<0.001

Welfare Analysis

- While the parameters of the model are of interest, analysts are often interested in measuring welfare implications from policy changes
- In general, the compensating variation associated with a change in site conditions or availability is implicitly defined by

$$\underset{\omega \in \tilde{A}^0}{Max}\left\{V_\omega(p_\omega^0, y; q^0, \gamma, \varepsilon)\right\}$$

$$= \underset{\omega \in \tilde{A}^1}{Max}\left\{V_\omega(p_\omega^1, y + C(p^0, q^0, p^1, q^1, y; \gamma, \varepsilon); q^1, \gamma, \varepsilon)\right\}$$

Welfare Analysis - Difficulties

1. For any given ε and γ, $C(p^0, q^0, p^1, q^1, y; \gamma, \varepsilon)$ is an implicit function for which no closed-form solution exists

   Solution: Numerical Bisection

2. Given $C(p^0, q^0, p^1, q^1, y; \gamma, \varepsilon)$, no closed form exists for

$$\bar{C}(p^0, q^0, p^1, q^1, y; \gamma) = E_\varepsilon\left[C(p^0, q^0, p^1, q^1, y; \gamma, \varepsilon)\right]$$

   Solution: Integration by simulation

3. Given an algorithm for computing $\bar{C}(p^0, q^0, p^1, q^1, y; \gamma)$, the analyst needs to recognize that the parameters used in the analysis are only estimates

   Solution: Bootstrapping

Algorithm

1. Resample with replacement N observations from the original sample and estimate the model. Repeat this process N_γ times, yielding a total of N_γ parameter vectors

$$\gamma^{(i)},\ i = 1, \ldots, N_\gamma$$

2. For each $\gamma^{(i)}$ and each observation in the sample (i.e., n=1,…,N), use McFaddens GEV similator to generate $\varepsilon^{(ink)}$, k=1,…,N_ε. Use numerical bisection to solve

$$\underset{\omega \in \tilde{A}^0}{Max}\left\{V_\omega\left(p_\omega^0, y; q^0, \gamma^{(i)}, \varepsilon^{(ink)}\right)\right\}$$
$$= \underset{\omega \in \tilde{A}^1}{Max}\left\{V_\omega\left(p_\omega^1, y + C(p^0, q^0, p^1, q^1, y; \gamma^{(i)}, \varepsilon); q^1, \gamma^{(i)}, \varepsilon^{(ink)}\right)\right\}$$

Algorithm (cont'd)

3. Average $C^{(ink)}$ over the N_ε draws from the disturbance distribution and the N observations in the sample to yield $\hat{C}^{(i)}$ as an estimate of

$$E_\varepsilon\left[C(p^0, q^0, p^1, q^1, y; \gamma^{(i)}, \varepsilon)\right]$$

4. The distributional characteristics of $\hat{C}^{(i)}$ over the various draws of $\gamma^{(i)}$ reflect uncertainty in the estimated parameters.

$$\hat{\hat{C}} = \frac{1}{N_\gamma}\sum_{i=1}^{N_\gamma}\hat{C}^{(i)}$$

Welfare Analysis - Scenarios

- Scenario A: Loss of Lake Michigan Trout
  - achieved by eliminating artificial stocking programs
- Scenario B: Loss of Lake Michigan Coho Salmon
  - achieved by eliminating artificial stocking programs
- Scenario C: 20% reduction in Toxin Levels

Welfare Analysis - Results

Scenario	Mean Compensating Variation	
	EV Model	GEV Model
Scenario A: Loss of Lake Trout	15.97 (269.19)	-37.10 (272.78)
Scenario B: Loss of Coho Salmon	274.18 (123.18)	304.82 (192.20)
Scenario C: 20% Toxin Reduction	-89.35 (54.37)	-108.13 (51.73)

Advantages and Limitations of KT Model

- Primary advantage: provides a unified and utility theoretic model of
  - corners
  - interior levels for consumed commodities
- Limitations (to date)
  - Econometrically challenging to estimate, requiring multivariate integration
  - Welfare analysis is non-trivial
  - To date has been estimated only with relatively simple functional form specifications for utility.

The Dual to the Kuhn-Tucker Model

- Originally introduced by Lee and Pitt (1986) and applied by Phaneuf (1999).
- Focus is on indirect utility function

$$H(v;\theta,\varepsilon) = \underset{q}{Max}\left\{U(q;\theta,\varepsilon) \mid v'q = 1\right\}$$

where

Note lack of nonnegativity constraints

$U(q;\theta,\varepsilon)$	is strictly quasi-concave in q
$q = (q_1, \ldots, q_M)'$	denotes commodities being analyzed
$v = (v_1, \ldots, v_M)'$	denotes normalized prices
θ	denotes parameters to be estimates

Notional Demands

- The application of Roy's identity allows us to recover *notional demands*

$$q_i(v;\theta,\varepsilon) = \frac{\partial H(v;\theta,\varepsilon)}{\partial v_i}\left[\sum_{j=1}^{M} v_j \frac{\partial H(v;\theta,\varepsilon)}{\partial v_j}\right]^{-1} \qquad i = 1,\ldots,M$$

- The q_i's are *notional* because they can be negative, since the original maximization problem did not include nonnegativity constraints
- They can be viewed as *latent* variables, with the observed variables being actual demands.

The Link Between Latent and Actual Demands

- The link between latent and actual demands comes through the notion of *virtual* prices

virtual prices are the prices that rationalize the pattern of zero consumptions; i.e., if the first K commodities are not consumed, then the virtual prices $\pi = \left(\pi_1, \ldots, \pi_K\right)'$ solve

$$0 = \frac{\partial H\left(\pi_1\left(\overline{v};q,\varepsilon\right), \ldots, \pi_K\left(\overline{v};q,\varepsilon\right), \overline{v};\theta,\varepsilon\right)}{\partial v_i} \qquad i = 1, \ldots, K$$

where

$$\overline{v} = \left(v_{K+1}, \ldots, v_M\right)'$$

The Link Between Latent and Actual Demands (cont'd)

- Observed demand is then given by

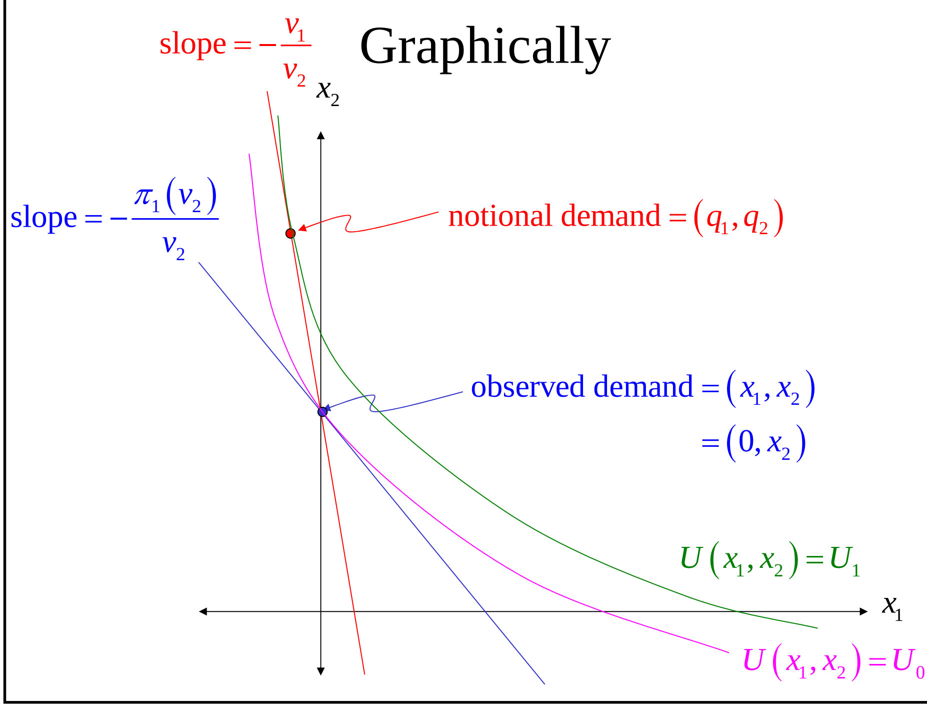
$$x_i\left(\tilde{v};\theta,\varepsilon\right) = \frac{\partial H\left(\tilde{v};\theta,\varepsilon\right)}{\partial v_i}\left[\sum_{j=1}^{M} \tilde{v}_j \frac{\partial H\left(\tilde{v};\theta,\varepsilon\right)}{\partial v_j}\right]^{-1} \qquad i = 1, \ldots, M$$

$$= \begin{cases} 0 & i = 1, \ldots, K \\ \dfrac{\partial H\left(\tilde{v};\theta,\varepsilon\right)}{\partial v_i}\left[\displaystyle\sum_{j=1}^{M} \tilde{v}_j \frac{\partial H\left(\tilde{v};\theta,\varepsilon\right)}{\partial v_j}\right]^{-1} & i = K+1, \ldots, M \end{cases}$$

where

$$\tilde{v} = \left(\pi\left(\overline{v};q,\varepsilon\right)', \overline{v}'\right)'$$

Graphically



Distributional Restrictions

- Notice that for the commodities that are not consumed

$$v_i \geq \pi_i(\overline{v}; q, \varepsilon)$$

This captures the distributional restrictions imposed by the boundary conditions

- Phaneuf (1999) uses a translog specification for the indirect utility function to model analyze recreational angling in the Wisconsin Great Lakes region