Limited Dependent Variables and Panel Data

„Logit, Probit and Friends“

Benjamin Bittschi

Sebastian Koch

Outline

- Binary dependent variables
  - Logit - Fixed Effects Models
  - Probit - Random Effects Models
- Censored dependent variables
- Empirical application – timber supply

Binary dependent variables

- Y_{it} =1: event happens
- Y_{it}=0: event does not happen

- P_{it}: i‘s probability for a 1 or zero at time t

- $P_{it} = Pr[y_{it}=1] = E(y_{it} | x_{it}) = F(x'_{it}ß)$

- Linear probability model: $F(x'_{it}ß) = x'_{it}ß$ or $F(w)=w$
  - > usual panel methods apply but $\hat{y}_{it}$ is not guaranteed to lie in the unit interval

Binary dependent variables II

- Solution to the “unity problem“:
  - Logistic or
  - Normal c.d.f

  that constrain F (w) between 0 and 1.

  - Logit $F(w) = \frac{e^w}{1+e^w}$

    or
  - Probit Model $F(w) = \int_{-\infty}^{w} \frac{1}{\sqrt{2\pi}} e^{\frac{-u^2}{2}} du = \phi(w)$

Binary dependent variables III

- Example: Neoclassical labour supply function

Worker participation in the labour force:

$y_{it} = 1$ if $y_{it}^* > 0$

$y_{it} = 0$ if $y_{it}^* \leqq 0$

(y_{it}^*: Δ offered wage / unobserved reservation wage)

where: $y_{it}^* = x'_{it}ß + u_{it}$, so that

$Pr[y_{it}=1] = Pr[y_{it}^*>0] = Pr[u_{it}>-x'_{it}ß] = F(x'_{it}ß)$

F.E. Model I

- Incidental parameters problem

F.E. Model: $y_{it}^* = x'_{it}ß+\mu_i+v_{it}$ with

$Pr[y_{it}=1] = Pr[y_{it}^*>0] = Pr[v_{it} > -x'_{it}ß-\mu_i] = F(x'_{it}ß+\mu_i)$

μ_i and ß unknown parameters and as N -> ∞ for fixed T, μ_i increase with N -> thus μ_i cannot be consistently estimated for fixed T -> incidental parameters problem

Neyman, Scott (1948)

Lancaster (2000)

F.E. Model II

- In the linear panel data model ß was estimated consistently by getting rid of μ_i -> within transformation

- Possible because mle of ß and μ_i are asymptotically independent
  => not longer the case for qualitative independent variables with fixed T

  - Hsiao (2003: 194f) shows for a simple model with N -> ∞ and T=2:

$$\text{plim } (N \to \infty)\ \hat{ß} = 2ß$$

  - Greene (2004) shows:
  - N=1000, T=2, 200 reps, 100% bias
  - N=1000, T=10, 200 reps, 16% bias
  - N=1000, T=20, 200 reps, 6.9% bias

F.E. Model III

Solution:

- Find a minimal sufficient statistic for μ_i
- For the logit model this is $\sum_{t=1}^{T} y_{it}$ according to Chamberlain (1980).
- Therefore: Maximizing conditional likelihood function:

$L_c = \prod_{i=1}^{N} \Pr(y_{i1},\ldots,y_{it} \mid \sum_{t=1}^{T} y_{it})$ -> conditional logit estimates for ß

For F.E. probit models:

• We cannot find simple functions for the parameters of interest that are independent of the parameters μ_i.

•There does **NOT** exist a consistent estimator of ß for the fixed effects probit models.

RE models

- R.E. probit model
- $u_{it} = \mu_i + v_{it}$ $\quad \left.\begin{array}{l} \mu_i \sim IIN(0,\sigma^2_\mu) \\ v_{it} \sim IIN(0,\sigma^2_v) \end{array}\right\}$ Independent of each other and of x_{it}

- Since $E(u_{it}u_{is}) = \sigma^2_\mu$ for $t \neq s$, the joint likelihood of $(y_{1t},\ldots,y_{Nt})$ can no longer be written as the product of the marginal likelihoods of the y_{it} .

- => complication of the derivation of maximum likelihood which will now involve T-dimensional integrals.

$$L_i = \Pr[y_{i1,} y_{i2},\ldots, y_{iT} / X] = \int\ldots\int f(u_{i1}, u_{i2},\ldots, u_{iT})du_{i1}du_{i2}\ldots du_{iT}$$

RE models II

- Maximization of L_i w.r.t. to ß and σ_μ infeasible if T is big.
- Trick: Write the joint density function as a product of the conditional density and the marginal density of μ_i.
- By conditioning on the individual effects, the T-dimensional integral reduces to:

$$f(u_{i1}, u_{i2}, \ldots, u_{iT}) = \int \prod_{t=1}^{T} f_1(u_{it} \mid \mu_i) f_2(\mu_i) d\mu_i$$

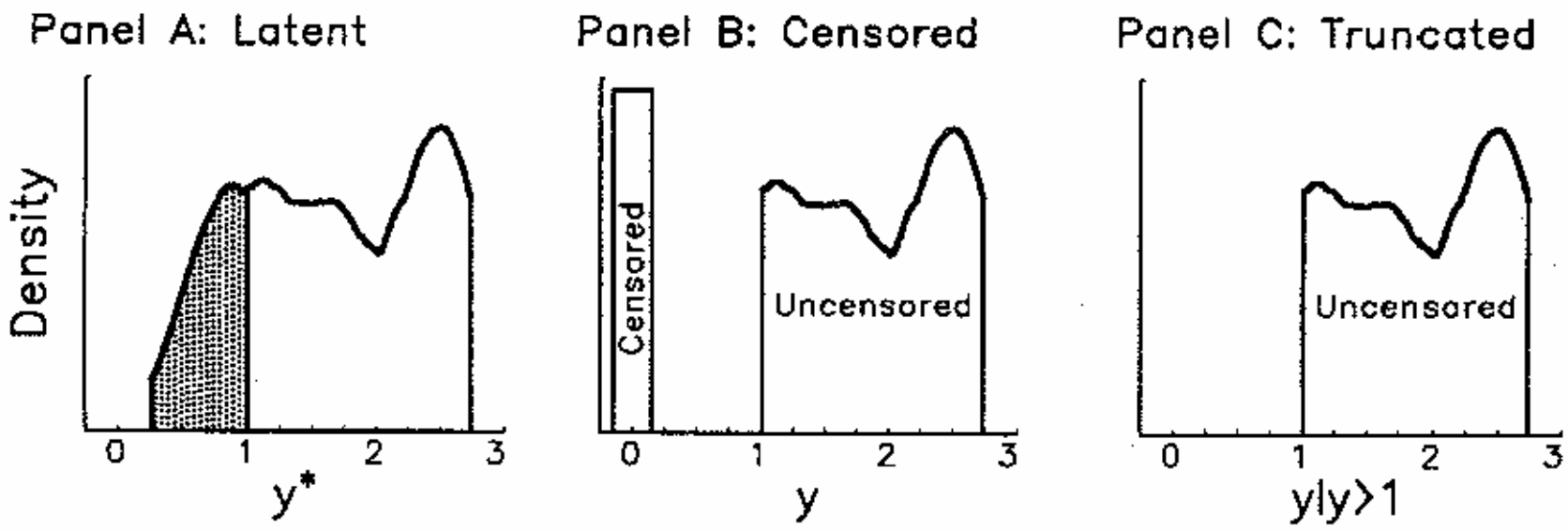
Part II: applied

But before:

- Tobit
- Censored
- Truncated
- Corner Solution
- Count Data
- Tobit II
- Hurdle model

Tobit Models:
Censored, truncated and corner solution models

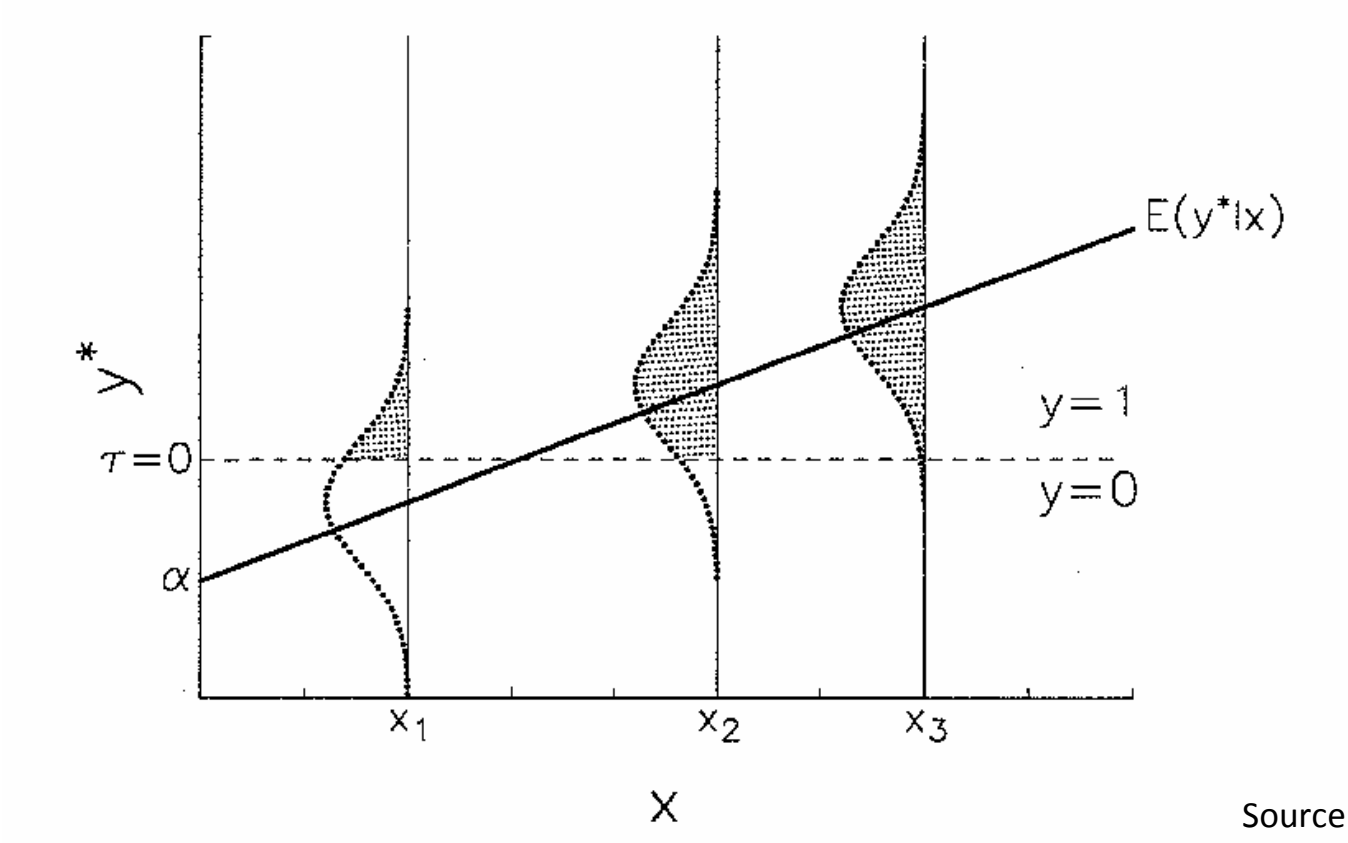
$$y_i = \begin{cases} y_i^* & \text{if } y_i^* > 1 \\ 0 & \text{if } y^* \leq 1 \end{cases}$$



Source: Long (1997)

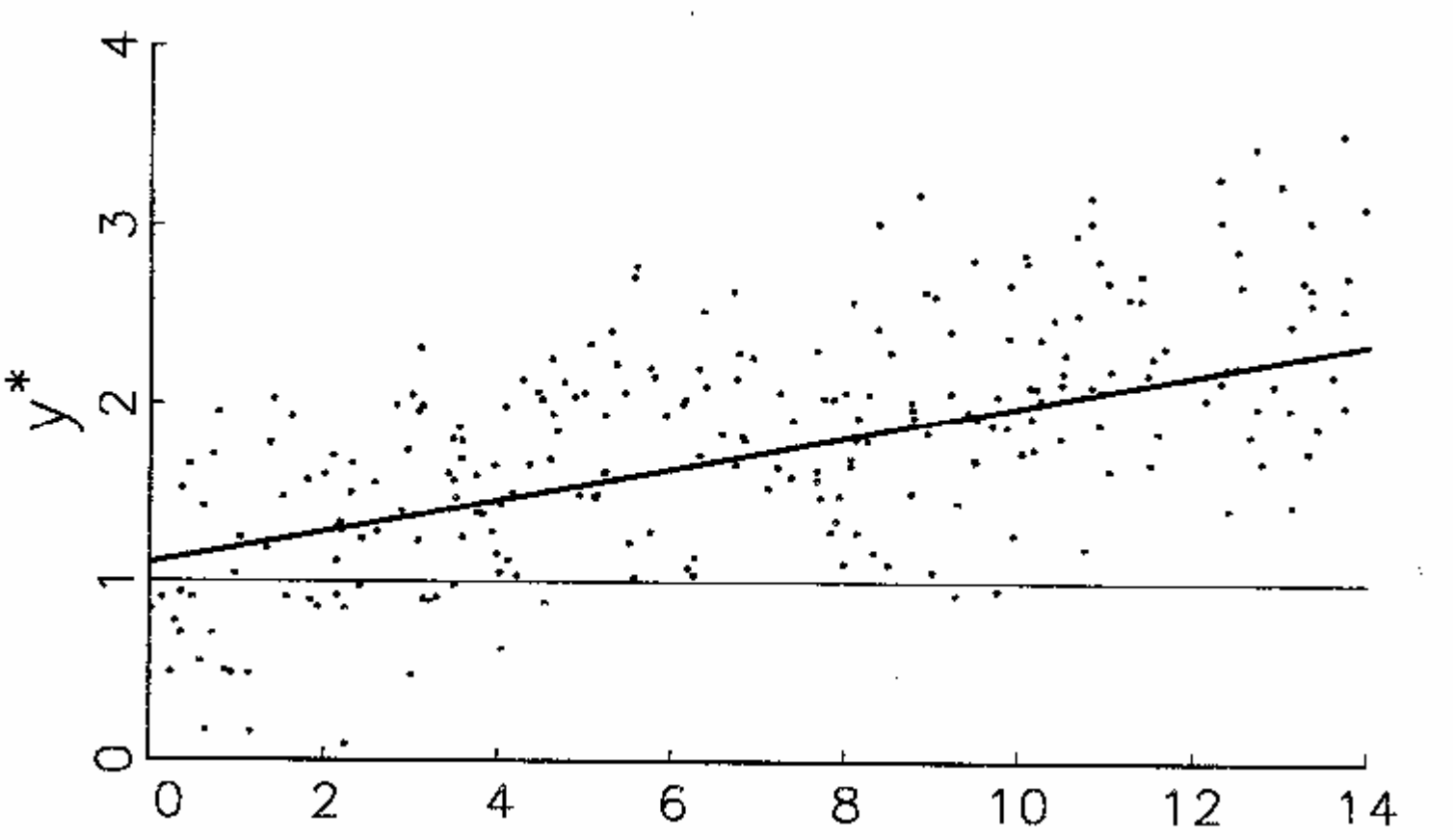
Concept of latent variable

- Distribution of y* given x in the binary response model



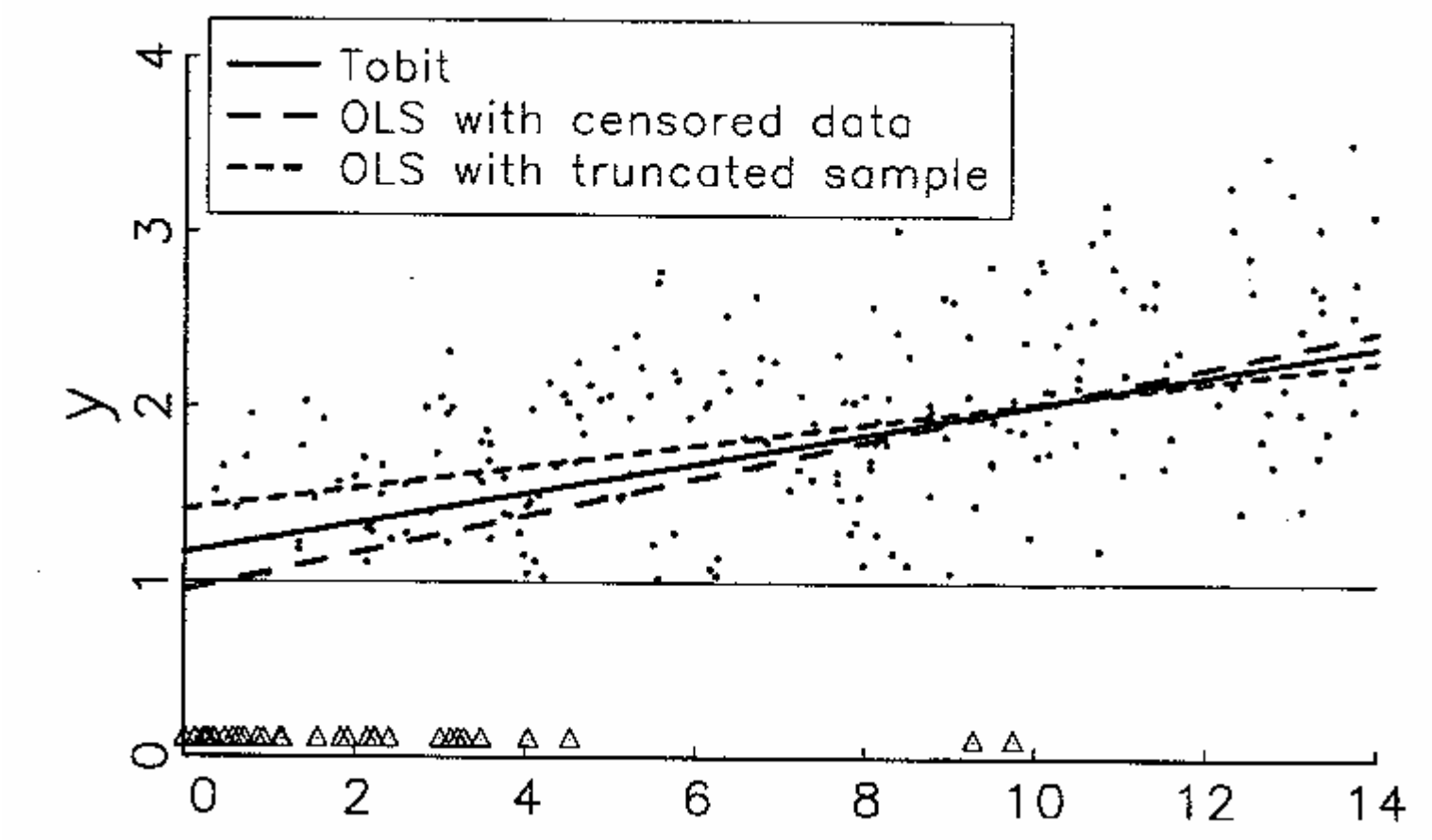
Source: Long (1997)

Regression without censoring



Source: Long (1997)

Regression with Censoring and Truncation



Source: Long (1997)

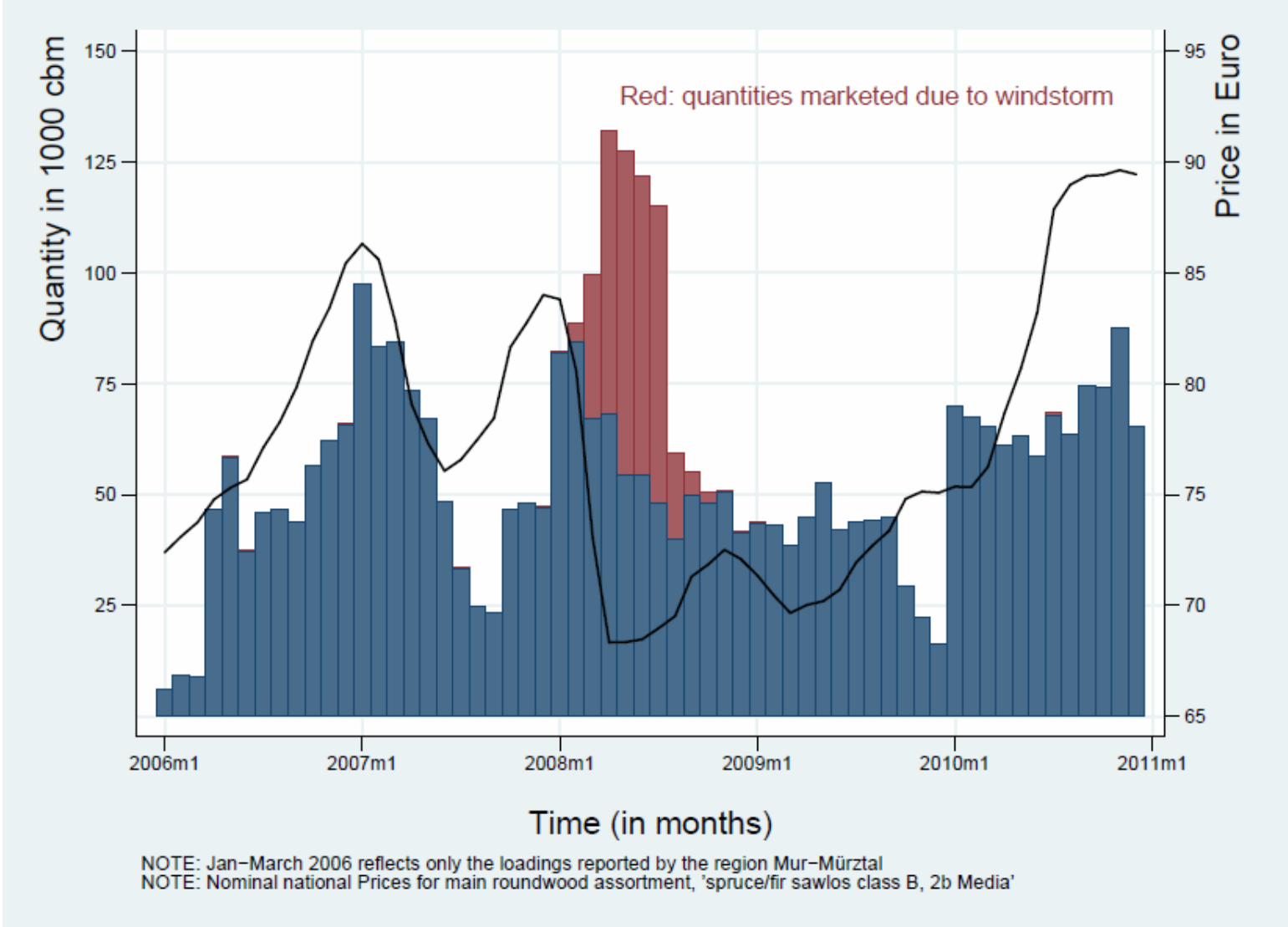
Application: Timber Supply

- T = 60 months
- N = 12.000 members
- → 720.000 observations (95% of obs.: supply = 0)

- Variable of interest:
  Timber supply

- Possibly influenced by:
  - –price (main assortment)
  - –wind storm (dummy)
  - –forest land (in ha)
  - –structure (agrar vs forest)
  - –sex, age of supplier
  - –seasons (dummy)
  - –etc. (not yet available)

Overview: aggreated monthly supply



Tobit and Tobit II

Tobit:

(Corner Solution Model ≈ censored model)

xttobit:

$$y_{it}^{*} = x_{it}'\beta + \mu_i + \nu_{it}$$

$$y_{it} = y_{it}^{*} \text{ if } y_{it}^{*} > 0$$
$$= 0 \quad \text{otherwise}$$

Interpretation?

-> coefficients of y* - not y

Tobit II:

(Two step or hurdle model)

$$y_{1it}^{*} = x_{1it}'\beta_1 + \mu_{1i} + \nu_{1it}$$
$$y_{2it}^{*} = x_{2it}'\beta_2 + \mu_{2i} + \nu_{2it}$$

xtlogit:
(or xtprobit)

$$y_{1it} = 1 \quad \text{if} \quad y_{1it}^{*} > 0$$
$$= 0 \quad \text{otherwise}$$

xtreg:

$$y_{2it} = y_{2it}^{*} \quad \text{if} \quad y_{1it} = 1$$

Interpretation?

-> latent variable, odds ratio, standard

Source: Baltagi (2008)

TOBIT: xttobit

```
. xttobit  lnqpm lnp ws lnfha sex age struct summer autumn winter, ll(0)

[iterations ommitted]

Random-effects tobit regression                 Number of obs      =    343140
Group variable: stataid                         Number of groups   =      5719

Random effects u_i ~ Gaussian                   Obs per group: min =        60
                                                               avg =      60.0
                                                               max =        60

                                                Wald chi2(9)       =   7951.60
Log likelihood  = -201025.89                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
       lnqpm |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lnp |   11.18171   .2172352    51.47   0.000     10.75594    11.60748
          ws |   2.795183   .0509047    54.91   0.000     2.695412    2.894954
       lnfha |    1.25643    .046126    27.24   0.000     1.166025    1.346835
         sex |  -.1124808   .0745891    -1.51   0.132    -.2586728    .0337111
         age |  -.0021527   .0026798    -0.80   0.422     -.007405    .0030997
      struct |   1.181999   .2358121     5.01   0.000     .7198163    1.644183
      summer |  -1.589878   .0379653   -41.88   0.000    -1.664289   -1.515467
      autumn |  -2.361483   .0424898   -55.58   0.000    -2.444761   -2.278205
      winter |   -1.40059    .044777   -31.28   0.000    -1.488351   -1.312829
       _cons |  -58.05237   .9729089   -59.67   0.000    -59.95924   -56.14551
-------------+----------------------------------------------------------------
    /sigma_u |   2.330702   .0288463    80.80   0.000     2.274165     2.38724
    /sigma_e |   5.027688   .0214695   234.18   0.000     4.985608    5.069767
-------------+----------------------------------------------------------------
         rho |   .1768872    .003482                      .1701459    .1837949
------------------------------------------------------------------------------

  Observation summary:    300751  left-censored observations
                           42389     uncensored observations
                               0 right-censored observations
```

Interpretation Tobit

Tobit:

(Corner solution model ≈ censored model)

Interpretation:

For a unit change in x_k , there is an expected change of $ß_k$ units in y*, holding all other variables constant.

In the above case:

A 1% increase in price, yields in an expected 11.2 % increase in the **propensity** to supply

(≠ Price elasticity of supply)

Source: Long (1997)

Tobit and Tobit II

Tobit:

(Corner Solution Model ≈ censored model)

xttobit:

$$y_{it}^{*} = x_{it}'\beta + \mu_i + \nu_{it}$$

$$y_{it} = y_{it}^{*} \text{ if } y_{it}^{*} > 0$$
$$= 0 \quad \text{otherwise}$$

Interpretation?

-> coefficients of y* - not y

Tobit II:

(Two step or hurdle model)

$$y_{1it}^{*} = x_{1it}'\beta_1 + \mu_{1i} + \nu_{1it}$$
$$y_{2it}^{*} = x_{2it}'\beta_2 + \mu_{2i} + \nu_{2it}$$

xtlogit:
(or xtprobit)

$$y_{1it} = 1 \quad \text{if} \quad y_{1it}^{*} > 0$$
$$= 0 \quad \text{otherwise}$$

xtreg:

$$y_{2it} = y_{2it}^{*} \quad \text{if} \quad y_{1it} = 1$$

Interpretation?

-> latent variable, odds ratio, standard

Source: Baltagi (2008)

TOBIT II: xtlogit

```
xtlogit yn price ws fha sex age struct summer autumn winter

[iterations ommitted]

Random-effects logistic regression              Number of obs      =    343140
Group variable: stataid                         Number of groups   =      5719

Random effects u_i ~ Gaussian                   Obs per group: min =        60
                                                               avg =      60.0
                                                               max =        60

                                                Wald chi2(9)       =   7508.92
Log likelihood  = -111755.31                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
          yn |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       price |   .0521769   .0010616    49.15   0.000     .0500961    .0542576
          ws |   1.035346   .0189913    54.52   0.000     .9981242    1.072569
         fha |   2.86e-06    .000061     0.05   0.963    -.0001166    .0001223
         sex |  -.0253376   .0313188    -0.81   0.419    -.0867212     .036046
         age |  -.0003606   .0011255    -0.32   0.749    -.0025665    .0018453
      struct |   2.040648   .0739547    27.59   0.000     1.895699    2.185596
      summer |   -.593911   .0146278   -40.60   0.000    -.6225809    -.565241
      autumn |  -.9269553   .0166207   -55.77   0.000    -.9595313   -.8943793
      winter |  -.5654742    .017482   -32.35   0.000    -.5997384   -.5312101
       _cons |  -7.094474   .1152076   -61.58   0.000    -7.320277   -6.868671
-------------+----------------------------------------------------------------
    /lnsig2u |   -.042231   .0243089                     -.0898756    .0054136
-------------+----------------------------------------------------------------
     sigma_u |   .9791059   .0119005                      .956057     1.00271
         rho |   .2256431   .0042475                     .2174272    .2340766
------------------------------------------------------------------------------
Likelihood-ratio test of rho=0: chibar2(01) =  2.1e+04 Prob >= chibar2 = 0.000
```

TOBIT II: xtlogit, odds ratio

```
xtlogit yn price ws fha sex age struct summer autumn winter, or

[iterations omitted]

Random-effects logistic regression              Number of obs      =    343140
Group variable: stataid                         Number of groups   =      5719

Random effects u_i ~ Gaussian                   Obs per group: min =        60
                                                               avg =      60.0
                                                               max =        60

                                                Wald chi2(9)       =   7508.92
Log likelihood  = -111755.31                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
          yn |         OR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       price |   1.053562   .0011185    49.15   0.000     1.051372    1.055757
          ws |   2.816082    .053481    54.52   0.000     2.713188    2.922878
         fha |   1.000003    .000061     0.05   0.963     .9998834    1.000122
         sex |   .9749807   .0305352    -0.81   0.419     .9169327    1.036704
         age |   .9996395   .0011251    -0.32   0.749     .9974368    1.001847
      struct |   7.695591   .5691255    27.59   0.000       6.6572    8.895951
      summer |   .5521635   .0080769   -40.60   0.000     .5365578    .5682232
      autumn |   .3957568   .0065778   -55.77   0.000     .3830724    .4088613
      winter |   .5680907   .0099314   -32.35   0.000     .5489552    .5878931
-------------+----------------------------------------------------------------
    /lnsig2u |   -.042231   .0243089                     -.0898756    .0054136
-------------+----------------------------------------------------------------
     sigma_u |   .9791059   .0119005                      .956057     1.00271
         rho |   .2256431   .0042475                     .2174272    .2340766
------------------------------------------------------------------------------
Likelihood-ratio test of rho=0: chibar2(01) =  2.1e+04 Prob >= chibar2 = 0.000
```

Tobit II: Interpretation:

xtlogit:

For a unit increase in x_k , there is an expected logit change by $ß_k$, holding all other variables constant.

-> Meaning not intuitive. Better:

Odds Ratio: $e^{(ßk * delta)}$

For each additional unit (delta =1) in x_k , the odds are expected to change by a factor of $e^{(ßk * delta)}$, holding all other variables constant.

I.e. in the above case:

An increase in price by 10 Euro (delta =10) , increases the odds of becoming a supplier by a factor of 1.69. (= 1.054^{10} when looking at the xtlogit odds ratio coefficents or exp(0.052 *10) when looking at the xtlogit output.)

xtreg:

xtreg interpretation is straight forward. Be aware that the regression only includes those forest owners that indeed do supply. (reduced sample).

Source: Long(2008)

TOBIT II: xtreg (re)

```
. xtreg  lnqpm lnp ws lnfha sex age struct summer autumn winter if yn ==1

Random-effects GLS regression                   Number of obs      =     42389
Group variable: stataid                         Number of groups   =      5719

R-sq:  within  = 0.0372                         Obs per group: min =         1
       between = 0.2166                                        avg =       7.4
       overall = 0.1523                                        max =        50

Random effects u_i ~ Gaussian                   Wald chi2(9)       =   3337.50
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =    0.0000

------------------------------------------------------------------------------
       lnqpm |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lnp |   1.332235   .0637507    20.90   0.000     1.207286    1.457184
          ws |   .3355431   .0138382    24.25   0.000     .3084208    .3626655
       lnfha |   .3273657   .0108706    30.11   0.000     .3060598    .3486717
         sex |   .0054594   .0172261     0.32   0.751    -.0283032     .039222
         age |  -.0014195   .0006205    -2.29   0.022    -.0026356   -.0002034
      struct |   .1377088   .0553097     2.49   0.013     .0293038    .2461139
      summer |  -.2016421   .0107977   -18.67   0.000    -.2228052   -.1804791
      autumn |   .0480821    .012617     3.81   0.000     .0233532     .072811
      winter |   .1729488   .0129368    13.37   0.000     .1475932    .1983045
       _cons |  -3.437619    .281763   -12.20   0.000    -3.989864   -2.885374
-------------+----------------------------------------------------------------
     sigma_u |  .46304362
     sigma_e |  .81661611
         rho |  .24329557   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

TOBIT II : xtreg (re)

case 2: forest land < 50 ha

```
. xtreg  lnqpm lnp ws lnfha sex age struct summer autumn winter if yn ==1 & fha
< 50

Random-effects GLS regression                   Number of obs      =     33709
Group variable: stataid                         Number of groups   =      5092

R-sq:  within  = 0.0449                         Obs per group: min =         1
       between = 0.1595                                        avg =       6.6
       overall = 0.1066                                        max =        41

Random effects u_i ~ Gaussian                   Wald chi2(9)       =   2450.12
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =    0.0000

------------------------------------------------------------------------------
       lnqpm |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lnp |   1.532198   .0735463    20.83   0.000      1.38805    1.676346
          ws |   .3582413   .0154119    23.24   0.000     .3280346     .388448
       lnfha |   .3294826   .0135288    24.35   0.000     .3029666    .3559987
         sex |   .0070948   .0188282     0.38   0.706    -.0298078    .0439974
         age |  -.0014982   .0006791    -2.21   0.027    -.0028293   -.0001671
      struct |   .1337394   .0596049     2.24   0.025      .016916    .2505628
      summer |  -.2147239   .0121944   -17.61   0.000    -.2386244   -.1908233
      autumn |   .0612261   .0146212     4.19   0.000     .0325691     .089883
      winter |    .199711   .0145692    13.71   0.000     .1711559    .2282662
       _cons |  -4.317427   .3249442   -13.29   0.000    -4.954306   -3.680548
-------------+----------------------------------------------------------------
     sigma_u |  .47612762
     sigma_e |  .81709227
         rho |    .253481   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Thanks!

BACKUP SLIDES

TOBIT II: xtprobit

```
 xtprobit yn price ws fha sex age struct summer autumn winter

[iterations ommited]

Random-effects probit regression                Number of obs      =    343140
Group variable: stataid                         Number of groups   =      5719

Random effects u_i ~ Gaussian                   Obs per group: min =        60
                                                               avg =      60.0
                                                               max =        60

                                                Wald chi2(9)       =   7714.50
Log likelihood  = -111789.99                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
          yn |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       price |    .0285288   .0005682    50.21   0.000     .0274151    .0296426
          ws |    .5604678   .0104794    53.48   0.000     .5399285    .5810071
         fha |    3.54e-06   .0000316     0.11   0.911    -.0000584    .0000655
         sex |   -.0142358   .0162565    -0.88   0.381    -.0460979    .0176263
         age |   -.0002047    .000584    -0.35   0.726    -.0013493    .0009399
      struct |    1.079983   .0383025    28.20   0.000     1.004912    1.155055
      summer |   -.3283732   .0079037   -41.55   0.000    -.3438642   -.3128822
      autumn |   -.5066704   .0087862   -57.67   0.000     -.523891   -.4894498
      winter |   -.3129585   .0093929   -33.32   0.000    -.3313682   -.2945488
       _cons |   -3.898276   .0606617   -64.26   0.000    -4.017171   -3.779381
-------------+----------------------------------------------------------------
    /lnsig2u |   -1.330694   .0233557                     -1.376471   -1.284918
-------------+----------------------------------------------------------------
     sigma_u |     .514095   .0060035                      .5024619    .5259974
         rho |    .2090445   .0038618                      .2015764    .2167142
------------------------------------------------------------------------------
Likelihood-ratio test of rho=0: chibar2(01) =  2.1e+04 Prob >= chibar2 = 0.000
```

Rho

The output includes the additional panel-level variance component, which is parameterized as the log of the variance $\ln(\sigma_\nu^2)$ (labeled `lnsig2u` in the output). The standard deviation σ_ν is also included in the output (labeled `sigma_u`) together with ρ (labeled `rho`), where

$$\rho = \frac{\sigma_\nu^2}{\sigma_\nu^2 + 1}$$

which is the proportion of the total variance contributed by the panel-level variance component.

When `rho` is zero, the panel-level variance component is unimportant, and the panel estimator is not different from the pooled estimator. A likelihood-ratio test of this is included at the bottom of the output. This test formally compares the pooled estimator (probit) with the panel estimator.

Source: Stata help

1. $\theta = 0$, the transformation is **I**, RE-GLS becomes OLS. This happens if $\sigma_\mu^2 = 0$, no 'effects';
2. $\theta = 1$, the transformation is **Q**, RE-GLS becomes FE-LSDV. This happens if T is very large or $\sigma_\nu^2 = 0$ or σ_μ^2 is large;
3. The transformation can never become just **P**, which would yield the *between* estimator.

Source: Kunst