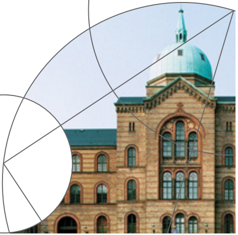
Censored Response: The Tobit Model

Advanced Microeconometrics

Anders Munk-Nielsen

2022



Plan for lectures: Helicopter

Part I: Linear methods. ✓

Part II: High-dimensional methods. ✓

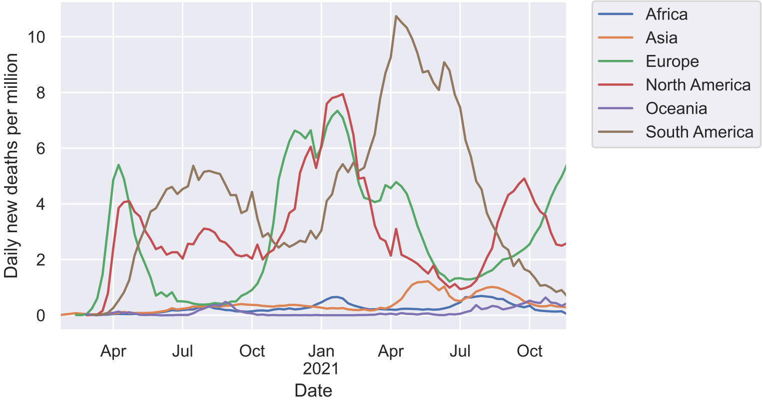
Part III: M-estimation, theory ✓

Part IV: M-estimation, concrete models ←

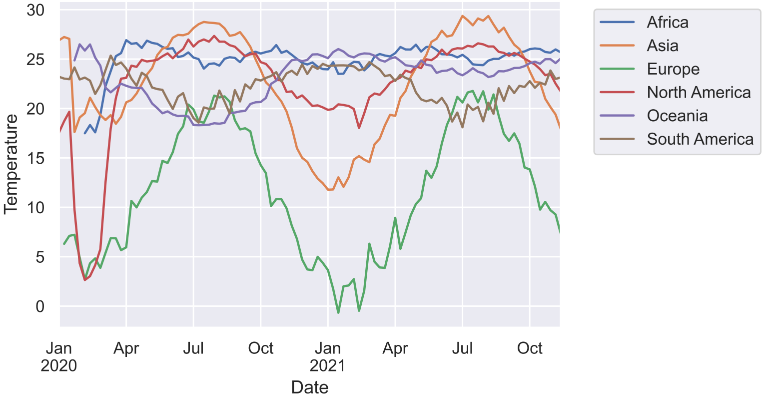
Where are we in the course?

Part	Topic	Parameterization non-linear	Estimation non-linear	Dimension dim(x)	Numerical optimization	M-estimation (Part III)	Outcome (y_i)	Panel (c_i)
I	OLS	$\div$	$\div$	low	$\div$	✓	$\mathbb{R}$	✓
II	LASSO	$\div$	✓	high	✓	$\div$	$\mathbb{R}$	$\div$
IV	Probit	✓	✓	low	✓	✓	$\{0,1\}$	$\div$
	Logit	✓	✓	low	✓	✓	$\{1,2,...,J\}$	$\div$
	Tobit	✓	✓	low	✓	✓	$[0;\infty)$	$\div$
	Sample selection	✓	✓	low	✓	✓	$\mathbb{R}$ and $\{0,1\}$	$\div$
	Simulated Likelihood	✓	✓	low	✓	✓	Any	✓
	Quantile Regression	$\div$	✓	(low)	✓	✓	$\mathbb{R}$	$\div$
	Non-parametric	✓	(✓)	∞	$\div$	$\div$	$\mathbb{R}$	$\div$

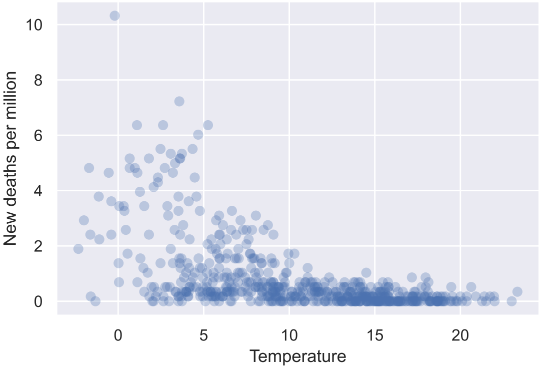
Covid Deaths Over Time



Temperature Matters



Temperature and Deaths? Data for Denmark



Discussion

1. What is the marginal effect of temperature on (expected) deaths?
2. How do we estimate it with OLS?

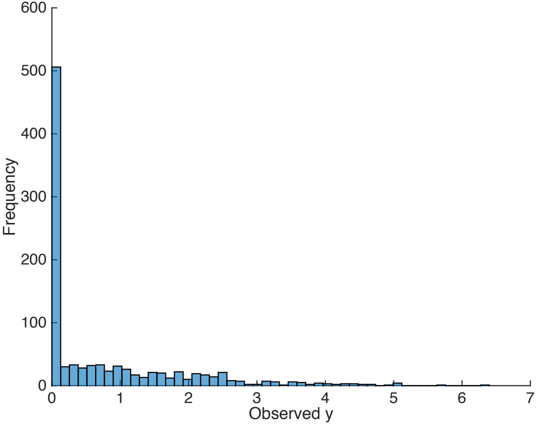
Tobit

1. Relevant empirical questions
   - Distinction: censoring or corner solutions.
2. The data generating process (the model)
3. The criterion function (deriving the likelihood function)
4. Features of interest: $\mathbb{E}(y|x)$ and $\mathbb{E}(y|\mathbf{x}, y > 0)$.
5. Specific issues.

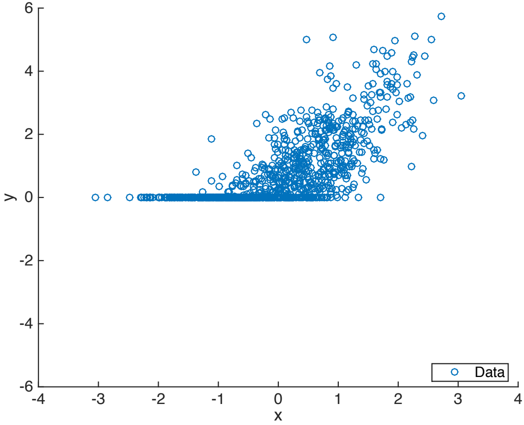
Agenda

Outline

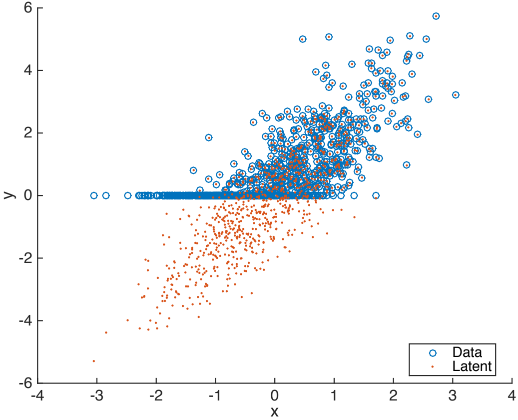
Data: Observed y



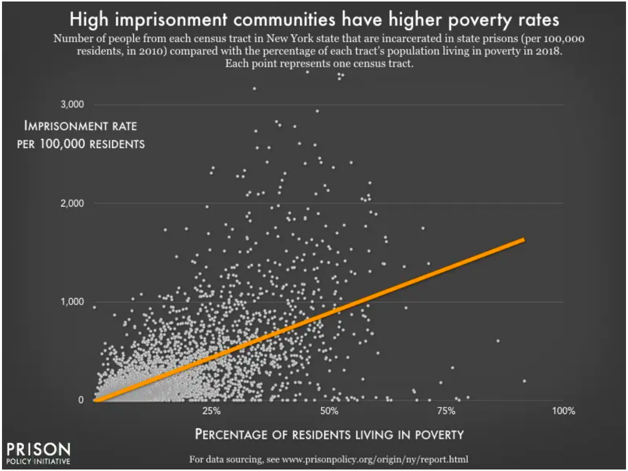
Data: y and x



Data: The latent model we will form

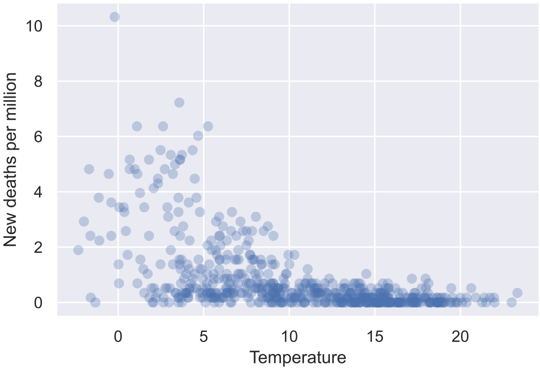


Crime Example

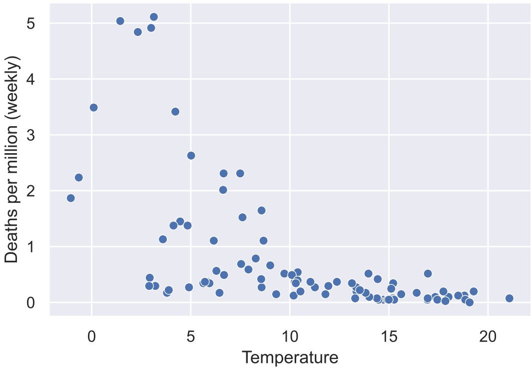


Source: prisonpolicy.org

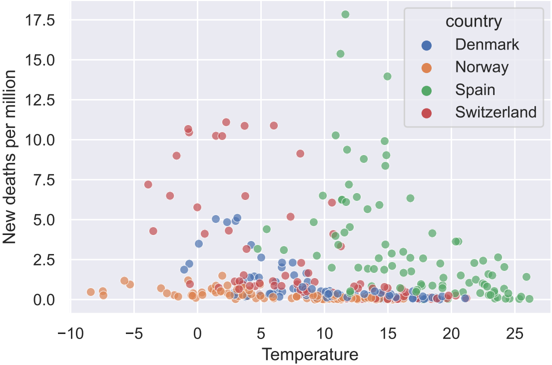
Covid Example, DK



Covid Example, DK: Weekly Level



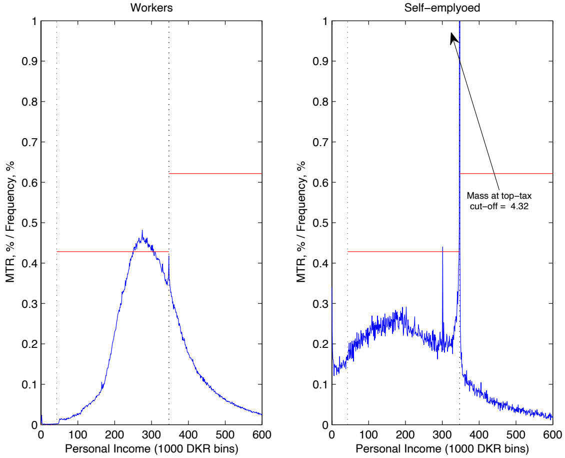
Covid Example



Examples

- **Hours worked.**
- **House valuation** must be positive.
- **Demand/consumption** cannot be negative.
- **Firm's exports:** Fixed costs of becoming an exporter
  - $\Rightarrow$ corner solution at zero.
- **Net wealth** (typically) cannot be negative.

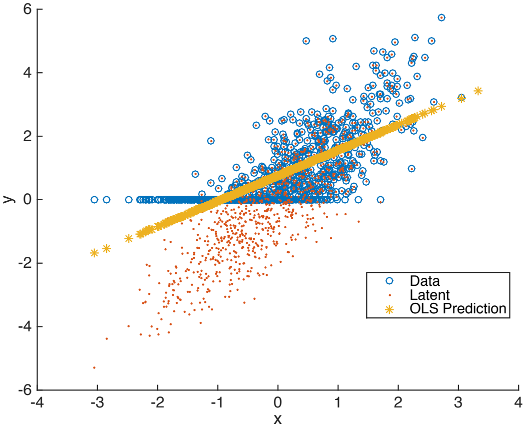
Example: Self-employed (2009)



Why censoring?

1. **Corner solution:** A feature of the economic problem.
   - E.g. non-negative consumption/investment/...
   - **Not interested** in the "latent, uncensored" variable.
2. **Censoring:** An issue with the data.
   - E.g. top-coding of wealth or earnings due to anonymity; ...
   - **Interested** in the latent, true data variable.

What OLS would do

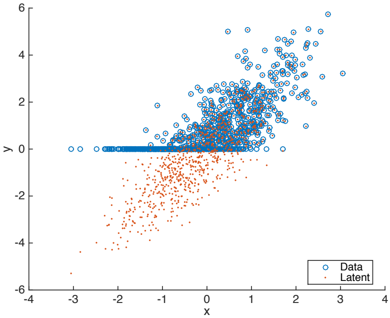


Goal: Get a more interesting answer for **the effect of** x **on** y.

Outline

Tobit Model

$$y_i^* = \mathbf{x}_i \beta_o + \varepsilon_i, \quad \varepsilon_i \sim \text{IID}\mathcal{N}(0, \sigma^2),$$
$$y_i = \max(y_i^*, 0).$$



Model

Tobit Model

$$
\begin{aligned}
y_i^* &= \mathbf{x}_i \beta_o + \varepsilon_i, \quad \varepsilon_i \sim \text{IID}\mathcal{N}(0, \sigma^2), \\
y_i &= \max(y_i^*, 0).
\end{aligned}
$$

Discussion

Here, σ is identified. Explain why (intuitively).

Hint: The latent variable, y_i^*, is *partially* observed.

Model in Python

Tobit Model

$$
\begin{aligned}
y_i^* &= \mathbf{x}_i \beta_o + \varepsilon_i, \quad \varepsilon_i \sim \text{IID}\mathcal{N}(0, \sigma^2), \\
y_i &= \max(y_i^*, 0).
\end{aligned}
$$

```python
def sim_data(theta,N):
    sig = theta[-1] # last element is sigma (not sigma^2)
    b = theta[:-1]  # ... beta is before
    K = b.size

    xx = np.random.normal(size=(N,K-1)) # continuous regressors
    oo = np.ones((N,1)) # constant
    x  = np.hstack([oo,xx])

    u = sig * np.random.normal(size=N)
    ystar = x@b + u # unobserved, so we do not return it
    y = np.fmax(ystar, 0.0) # elementwise max
    return y,x
```

Outline

Deriving the Density

- **Challenge:** We have assumptions on y_i^*, not y_i...
- **Solution:** "elephant strategy",

$$f(y_i|x_i) = \begin{cases} f_1(y_i|\mathbf{x}_i) & \text{if } y_i < 0, \\ f_2(y_i|\mathbf{x}_i) & \text{if } y_i = 0, \\ f_3(y_i|\mathbf{x}_i) & \text{if } y_i > 0. \end{cases}$$

  - we can always split a function up like this.

- **Goal:** Derive $f(y_i|x_i)$.
  - **Method:** Split into 3 cases;

1. $y_i < 0$**:** This can never happen, so

$$f(y_i|x_i) = 0 \text{ for all } y_i < 0.$$

2. $y_i = 0$**:** Happens if $y_i^* \leq 0$. There will be a *mass point* in y_i at 0:

$$f(0|x_i) = \Pr(y_i^* \leq 0|x_i).$$

3. $y_i > 0$**:** Happens if $y_i^* > 0$.
Here, $y_i^* = y_i$ so the distributions are identical

$$f(y_i|\mathbf{x}_i) = f(y_i^*|\mathbf{x}_i) \text{ for all } y_i > 0.$$

Case 2: $y_i = 0$

- **Model:**

$$y_i^* = \mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i, \quad \varepsilon_i | x_i \sim \mathcal{N}(0, \sigma^2).$$

- **Proposition:** Show that

$$\Pr(y_i^* \leq 0) = 1 - \Phi(\mathbf{x}_i\boldsymbol{\beta}/\sigma),$$

where Φ and Φ are the CDF and PDF of the standard normal.

- **Proof:**

$$\begin{aligned}
\Pr(y_i^* \leq 0|x_i) &= \Pr(\varepsilon_i < -\mathbf{x}_i\boldsymbol{\beta}|\mathbf{x}_i) \\
&= \Phi(-\mathbf{x}_i\beta/\sigma) \\
&\overset{\text{sym.}}{=} 1 - \Phi(\mathbf{x}_i\beta/\sigma).
\end{aligned}$$

(symmetry: $F(a) = 1 - F(-a)$)

Case 3: $y_i > 0$

- **Model:**

$$y_i^* = \mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i, \quad \varepsilon_i | x_i \sim \mathcal{N}(0, \sigma^2).$$

- **Proposition:**

$$f(y_i|x_i) = \frac{1}{\sigma}\phi\left(\frac{y_i - \mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right) \text{ for all } y_i > 0.$$

- **Proof:** For all $y_i > 0$,

$$\begin{aligned} F(y_i|\mathbf{x}_i) &= \Pr(\mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i < y_i|\mathbf{x}_i) \\ &= \Phi\left(\frac{y_i - \mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right), \\ \Rightarrow f(y_i|\mathbf{x}_i) &= \phi\left(\frac{y_i - \mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right)\frac{1}{\sigma}. \end{aligned}$$

Putting the parts together

- **Collecting the terms:**

$$f(y_i|x_i) = \begin{cases} 0 & \text{if } y_i < 0, \\ 1 - \Phi(\mathbf{x}_i\beta/\sigma) & \text{if } y_i = 0, \\ \frac{1}{\sigma}\phi\left(\frac{y_i - \mathbf{x}_i\beta}{\sigma}\right) & \text{if } y_i > 0. \end{cases}$$

Tobit Loglikelihood

$$\ell_i(\theta) = \mathbf{1}_{\{y_i=0\}} \log\left[1 - \Phi\left(\frac{\mathbf{x}_i\beta}{\sigma}\right)\right] + \mathbf{1}_{\{y_i>0\}} \log\left[\frac{1}{\sigma}\phi\left(\frac{y_i - \mathbf{x}_i\beta}{\sigma}\right)\right].$$

Tobit Loglikelihood

$$\ell_i(\theta) = \mathbf{1}_{\{y_i=0\}} \log\left[1 - \Phi\left(\frac{\mathbf{x}_i\beta}{\sigma}\right)\right] + \mathbf{1}_{\{y_i>0\}} \log\left[\frac{1}{\sigma}\phi\left(\frac{y_i - \mathbf{x}_i\beta}{\sigma}\right)\right].$$

```python
def q(theta, y, x):
    return -loglikelihood(theta, y, x)

def loglikelihood(theta, y, x):
    sig = np.abs(theta[-1]) # ensure positivity
    b = theta[:-1]
    N,K = x.shape

    Phi = norm.cdf(xb@b / sig)
    phi = 1/sig * norm.pdf((y - x@b)/sig)

    ll = (y == 0.0) * np.log(1.0-Phi) + (y > 0) * np.log(phi)
    return ll
```

Discussion

Discuss

Think of examples where identification might break down.

- If the `rank` condition fails.
- If for all i, the outcome is `censored (all outcomes at y=0)`
- ...

`proof of contradiction.. -> can we tell two alternatives apart?`

Outline

Features of Interest

- **Depends** on the reason for censoring...

Corner solution: Interested in $\mathbb{E}(y|x)$ and $\mathbb{E}(y|\mathbf{x}, y > 0)$, as well as $\Pr(y = 0|\mathbf{x})$.

Data censoring: Interested in $\mathbb{E}(y^*|\mathbf{x})$.

(and, most importantly, the partial effects of these objects!)

Data Censoring

- **Interest** is in y^*
  - ... censoring is just a data issue.
  - Perhaps due to anonymity.
- **Implication:** Since $\mathbb{E}(y^*|\mathbf{x}_i) = \mathbf{x}_i\boldsymbol{\beta}$,
  - Standard linear interpretation (once we have β)
  - (e.g. constant partial effects, ...)
- **Other feature:** The censoring probability.

Censoring probability

- **Question:** Show that

$$\Pr(y_i > 0|\mathbf{x}_i) = \Phi\left(\frac{\mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right).$$

- **Answer:**

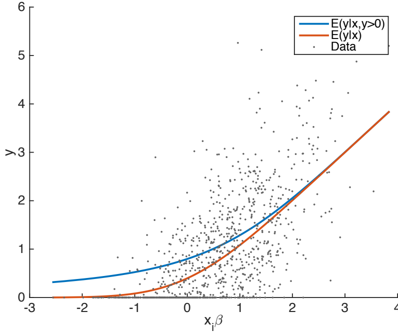
$$\begin{aligned}\Pr(y_i > 0|\mathbf{x}_i) &= 1 - \Pr(y_i \leq 0|\mathbf{x}_i) \\ &= 1 - \Pr(y_i = 0|\mathbf{x}_i) \\ &= \Phi\left(\frac{\mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right).\end{aligned}$$

Corner Solutions

- **Interest:** is now in actual y.
  suppose they do not work (wages are zero)
  - E.g. "what is the effect on wages of education".
- **Features of interest:** $\mathbb{E}(y|\mathbf{x})$ and $\mathbb{E}(y|\mathbf{x}, y > 0)$.
- **Challenge:** Non-linear objects.

Our goal

- **Goal:** Get realistic predictions for y_i depending on $\mathbf{x}_i\beta$.



out here, everyone is non-censored

Definition (truncated density)

$$f(a|A > k) = \mathbf{1}_{\{a>k\}} \frac{f(a)}{1 - F(k)}.$$

the truncated variable is proportional
to the untruncated var

▶▶ Derivation

- **Expectation:** for $A \sim \mathcal{N}(0, 1)$,

$$\begin{aligned}
\mathbb{E}(A|A > k) &= \int_k^\infty \frac{\phi(a)}{1 - \Phi(k)} a \, \mathrm{d}a \\
&= \frac{1}{1 - \Phi(k)} \int_k^\infty a\phi(a) \, \mathrm{d}a \\
&= \dots \\
&= \frac{1}{1 - \Phi(k)} \phi(k).
\end{aligned}$$

▶▶ Derivation

Conditional Expectation

- **Task:** Derive $\mathbb{E}(y_i|\mathbf{x}_i, y_i > 0)$.
  - **Note:** For all $y_i > 0$, $y_i = y_i^*$.
- **Algebra:**

$$
\begin{aligned}
\mathbb{E}(y_i|\mathbf{x_i}, y_i > 0) &= \mathbb{E}(y_i^*|\mathbf{x_i}, y_i^* > 0) \\
&= \mathbb{E}(\mathbf{x_i}'\beta + \varepsilon_i|\mathbf{x_i}, \mathbf{x_i}'\beta + \varepsilon_i > 0) \\
&= \mathbf{x_i}'\beta + \sigma\mathbb{E}\left(\frac{\varepsilon_i}{\sigma}\middle|\mathbf{x_i}, \frac{\varepsilon_i}{\sigma} > -\frac{\mathbf{x_i}'\beta}{\sigma}\right) \\
&= \mathbf{x_i}'\beta + \sigma\frac{\phi\left(-\frac{\mathbf{x_i}'\beta}{\sigma}\right)}{1 - \Phi\left(-\frac{\mathbf{x_i}'\beta}{\sigma}\right)} \\
&= \mathbf{x_i}'\beta + \sigma\lambda\left(\frac{\mathbf{x_i}'\beta}{\sigma}\right).
\end{aligned}
$$

- **Inverse Mill's Ratio:** $\lambda(z) \equiv \frac{\phi(z)}{\Phi(z)}$.
  - Interpretation: omitted regressor in a pure OLS (hence the bias).

Unconditional Expectation

- **Proposition:**

$$\mathbb{E}(y_i|\mathbf{x}_i) = \mathbf{x}_i\beta\Phi\left(\frac{\mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right) + \sigma\phi\left(\frac{\mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right).$$

- **Trick:** Note that

$$\begin{aligned}\mathbb{E}(y_i|\mathbf{x}_i) &= \Pr(y_i \leq 0|\mathbf{x}_i)\mathbb{E}(y_i|\mathbf{x}_i, y_i \leq 0) \\ &\quad + \Pr(y_i > 0|\mathbf{x}_i)\mathbb{E}(y_i|\mathbf{x}_i, y_i > 0).\end{aligned}$$

- **Proof:** Since $y_i \geq 0$,

$$\mathbb{E}(y_i|\mathbf{x}_i, y_i \leq 0) = \mathbb{E}(y_i|\mathbf{x}_i, y_i = 0) = 0.$$

Hence:

$$\begin{aligned}\mathbb{E}(y_i|\mathbf{x}_i) &= 0 + \Phi\left(\frac{\mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right)\left[\mathbf{x}_i\beta + \sigma\frac{\phi\left(\frac{\mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right)}{\Phi\left(\frac{\mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right)}\right], \\ &= \mathbf{x}_i\beta\Phi\left(\frac{\mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right) + \sigma\phi\left(\frac{\mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right).\end{aligned}$$

In Python

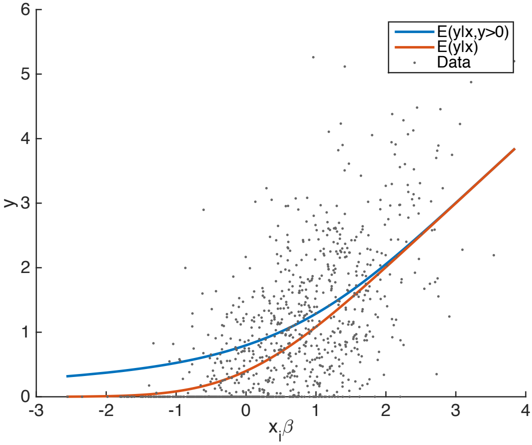
Features of interest

$$\begin{aligned}\mathbb{E}(y_i|\mathbf{x}_i) &= \mathbf{x}_i\beta\Phi\left(\frac{\mathbf{x}_i\beta}{\sigma}\right) + \sigma\phi\left(\frac{\mathbf{x}_i\beta}{\sigma}\right), \\ \mathbb{E}(y_i|\mathbf{x}_i, y_i > 0) &= \mathbf{x}_i\beta + \sigma\lambda\left(\frac{\mathbf{x}_i\beta}{\sigma}\right).\end{aligned}$$

```python
def mills_ratio(z):
    return norm.pdf(z) / norm.cdf(z)

def predict(theta, x):
    b = theta[:-1]
    s = theta[-1]
    xb = x@b
    E = xb * norm.cdf(xb/s) + s*norm.pdf(xb/s)
    Epos = xb + s*mills_ratio(xb/s)
    return E, Epos
```

Comparing



Properties

$$\mathbb{E}(y_i|\mathbf{x}_i) = \mathbf{x}_i\beta\Phi\left(\frac{\mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right) + \sigma\phi\left(\frac{\mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right),$$

$$\mathbb{E}(y_i|\mathbf{x}_i, y_i > 0) = \mathbf{x}_i\beta + \sigma\frac{\phi\left(\frac{\mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right)}{\Phi\left(\frac{\mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right)}.$$

Property 1. Both converge to $\mathbf{x}_i\boldsymbol{\beta}$.

Property 2. $\mathbb{E}(y_i|\mathbf{x}_i, y_i > 0) > \mathbb{E}(y_i|\mathbf{x}_i)$ for all x_i (when $\beta_p > 0 \forall p$).

Property 3. $\mathbb{E}\left[\max(y_i^*, 0)|\mathbf{x}_i\right] \geq \max\left[\mathbb{E}(y_i^*|\mathbf{x}_i), 0\right]$.
[Why?]

Partial Derivatives

Results

$$\begin{aligned}
\Pr(y_i > 0|\mathbf{x}_i) &= \Phi\left(\frac{\mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right), \\
\mathbb{E}(y_i|\mathbf{x}_i) &= x_i\boldsymbol{\beta}\Phi\left(\frac{\mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right) + \sigma\phi\left(\frac{\mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right), \\
\mathbb{E}(y_i|\mathbf{x}_i, y_i > 0) &= \mathbf{x}_i\beta + \sigma\frac{\phi\left(\frac{\mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right)}{\Phi\left(\frac{\mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right)}.
\end{aligned}$$

Derivatives

$$\left.\frac{\partial \Pr(y>0|\mathbf{x})}{\partial x_{ip}}\right|_{\mathbf{x}=\mathbf{x}^0} = \frac{\beta_p}{\sigma}\phi\left(\frac{\mathbf{x}^0\boldsymbol{\beta}}{\sigma}\right),$$

$$\left.\frac{\partial \mathbb{E}(y|\mathbf{x})}{\partial x_p}\right|_{\mathbf{x}=\mathbf{x}^0} = \beta_p\underbrace{\Phi\left(\frac{\mathbf{x}^0\boldsymbol{\beta}}{\sigma}\right)}_{\in(0;1)},$$

$$\left.\frac{\partial \mathbb{E}(y|\mathbf{x},y>0)}{\partial x_p}\right|_{\mathbf{x}=\mathbf{x}^0} = \beta_p\underbrace{\varphi\left(\frac{\mathbf{x}^0\boldsymbol{\beta}}{\sigma}\right)}_{\in(0;1)}$$

- **See graph:** They start flat, then converge to $\mathbf{x}_i\boldsymbol{\beta}$...
- (deriving the function $\varphi(\cdot)$ is involving but boatloads of fun)

Outline

Motivation: Heteroscedasticity

- **Suppose** the true model features heteroscedasticity. E.g.

Censoring and heteroscedasticity

$$\begin{aligned} y_i^* &= \mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i, \quad \varepsilon_i \sim \text{IID}\mathcal{N}\left(0, \exp(\mathbf{x}_i'\boldsymbol{\delta})\sigma^2\right), \\ y_i &= \max(y_i^*, 0). \end{aligned}$$

think the latent variale

- **Violates** distributional assumptions in Tobit...
  - [although one could just extend tobit and estimate δ in addition to β].

Weaker Assumptions

Censored Regression Model

$$
\begin{aligned}
y_i^* &= \mathbf{x}_i\beta + \varepsilon_i, \quad \text{Med}(\varepsilon_i|\mathbf{x}_i) = 0,\\
y_i &= \max(y_i^*, 0).
\end{aligned}
$$

- **Note:** Tobit is nested!

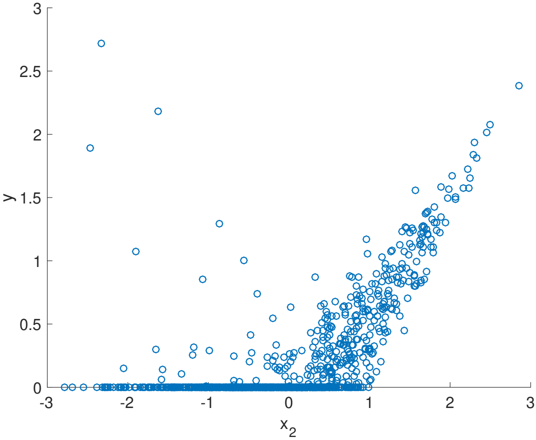
CLAD Criterion Function

CLAD Criterion

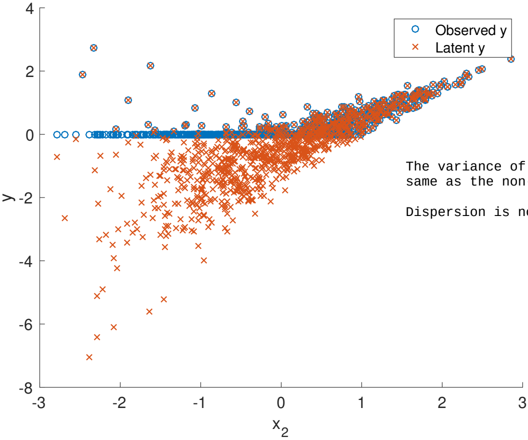
$$q(y_i, \mathbf{x}_i, \beta) = |y_i - \max(0, \mathbf{x}_i\beta)| \, .$$

- **Consistency?** It can be shown that the LAD works under the conditional *median* assumption from before...
  - ... just like LS works under a conditional *mean* assumption.
  - **Why?** details

Heteroscedasticity: Observed data



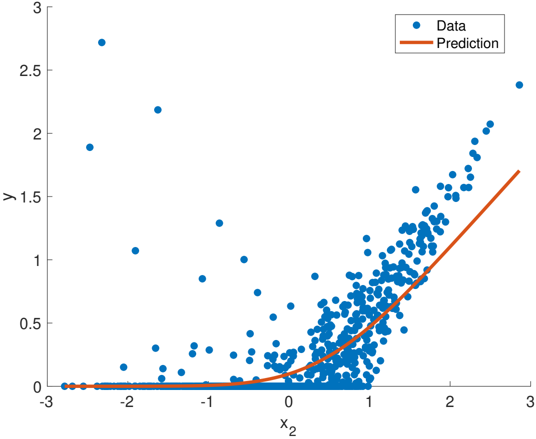
Heteroscedasticity: Latent index



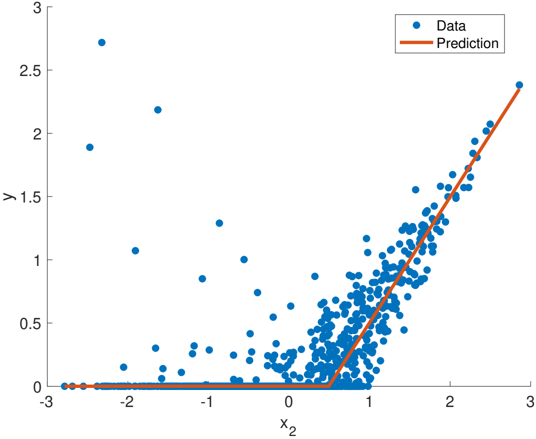

```
The variance of the censored data is not the
same as the non-censored data!!

Dispersion is not the same!
```

Tobit: Inconsistency



Clad: Consistency



Step back:

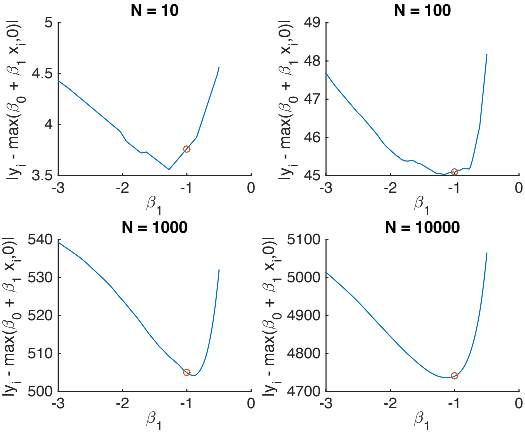
CLAD Criterion

very low likelihood that you find a y exactly at Max(0, x_i\beta)

$$q(y_i, \mathbf{x}_i, \beta) = |y_i - \max(0, \mathbf{x}_i\beta)| \, .$$

- **Smoothness fails:** recall what Jesper said
  - Non-differentiability at $\mathbf{x}_i\beta = 0$,
  - 2nd derivative is zero everywhere! -> inverse Hessian. Dividing by the inv(Hessian) then fails -> it's linear function around the kink
- **Magic:** $Q(\theta)$ becomes smooth as $N \to \infty$!
  [see graphs]
  - **Asymptotic normality?** Yes, but not from thm. 5.3.
- **Practical concern:** If $\mathbf{x}_i\beta < 0 \, \forall i$, the criterion function becomes flat!!

$N \to \infty$: Smoothness appears

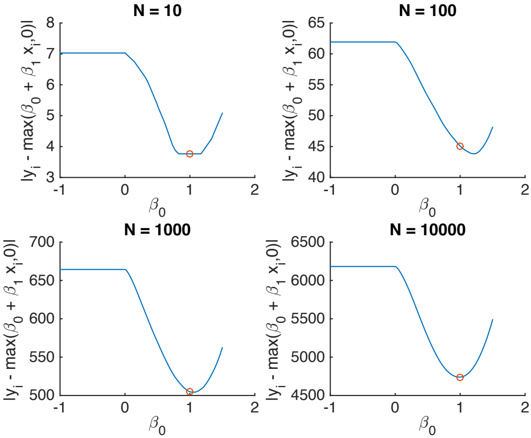


Smoothness: Even though $q_i(\beta) \equiv q(y_i, x_i, \beta)$ is non-smooth for all i, ...

When $N \to \infty$**:** $\sum_{i=1}^{N} q_i(\beta)$ becomes smooth anyway.

red dot shows the true value of -1

$N \to \infty$: Flatness persists



Flat region for β_0 low enough persists even as $N \to \infty$.
Starting optimizer here $\Rightarrow$ immediate termination.

Appendix: Truncated Density

- **Goal:** Show that
$$f(a|A > k) = \frac{f(a)}{1 - F(k)}.$$

- **First** we find the cdf.

$$\begin{aligned} F(a|A > k) &= \Pr(A < a|A > k) \\ &= \frac{\Pr(A < a \text{ and } A > k)}{\Pr(A > k)} \quad \text{(def. of cond. prob.)} \\ &= \frac{F(a) - F(k)}{1 - F(k)}. \end{aligned}$$

- **Thus:**
$$f(a|A > k) = \frac{\mathrm{d}}{\mathrm{d}a} F(a|A > k) = \frac{f(a)}{1 - F(k)}. \quad \blacksquare$$

◂ Back

Appendix: Truncated Expectation

Goal: Show that when $A \sim \mathcal{N}(0,1)$, $\mathbb{E}(A|A > k) = \frac{\phi(k)}{1-\Phi(k)}$.

$$
\begin{aligned}
\mathbb{E}(A|A > k) &= \int_k^\infty a\, f_{A|A>k}(a)\,\mathrm{d}a \\
&= \int_k^\infty a \frac{\phi(a)}{1-\Phi(k)}\,\mathrm{d}a \\
&= \frac{1}{1-\Phi(k)} \int_k^\infty -\frac{\mathrm{d}}{\mathrm{d}a}\phi(a)\mathrm{d}a \quad \left(\phi'(a) = -a\phi(a)\right) \\
&= \frac{1}{1-\Phi(k)} \int_{-\infty}^k \frac{\mathrm{d}}{\mathrm{d}a}\phi(a)\mathrm{d}a \quad \left(\text{"}\textstyle\int_a^b f = \int_b^a -f\text{"}\right) \\
&= \frac{1}{1-\Phi(k)} \frac{\mathrm{d}}{\mathrm{d}k} \int_{-\infty}^k \phi(a)\mathrm{d}a \\
&= \frac{\phi(k)}{1-\Phi(k)}. \quad \blacksquare
\end{aligned}
$$

◂ Back

Why a median assumption?

Property

For any non-decreasing function, g,

$$\text{Med}[g(y)] = g[\text{Med}(y)].$$

- **Implication:** With $z \mapsto \max(z, 0)$ as g,

$$\begin{aligned}
\text{Med}(y_i|\mathbf{x}_i) &= \text{Med}\left[\max(y_i^*, 0)\middle|\mathbf{x}_i\right] \\
&= \max\left[\text{Med}(y_i^*|\mathbf{x}_i), 0\right] \\
&= \max\left[\mathbf{x}_i\beta + \text{Med}(\varepsilon_i|\mathbf{x}_i), 0\right] \\
&= \max(\mathbf{x}_i\beta, 0).
\end{aligned}$$

  - **This** is the intuition for why we work with a median restriction.
- **Note:** We have not (and will not in this course) show *consistency*.

◂ Back