Part 2: Exam Project

Character count is around 13.000 without tables

12. januar 2025

Question 1

The conditional distribution of latent variable Y_i^* is Gaussian as the only stochastic variable is U_i which is normally distributed. Thus, we derive expressions for the conditional mean and variance of Y_i^* conditionally on X_i.
The conditional mean is:

$$\begin{aligned} E\left[Y_i^*|X_i\right] &= E\left[\alpha_0 + \beta_0 X_i + U_i|X_i\right] = \alpha_0 + \beta_0 X_i + E\left[U_i|X_i\right] \\ &= \alpha_0 + \beta_0 X_i + \mu_0 \end{aligned}$$

Using the conditional mean, the variance can be derived as:

$$\begin{aligned} E\left[\left(Y_i^* - E\left[Y_i^*|X_i\right]\right)^2 \middle| X_i\right] &= E\left[\left((\alpha_0 + \beta_0 X_i + U_i) - (\alpha_0 + \beta_0 X_i + \mu_0)\right)^2 \middle| X_i\right] \\ &= E\left[(U_i - \mu_0)^2 \middle| X_i\right] = E\left[\left(U_i - E\left[U_i|X_i\right]\right)^2 \middle| X_i\right] \\ &= \sigma_0^2 \end{aligned}$$

Thus, we now know that $Y_i^* \mid X_i = x \overset{d}{\sim} N(\alpha_0 + \beta_0 x + \mu_0, \sigma_0^2)$. Consequently, by shifting and rescaling appropriately, we can write the CDF with a standard normal:

$$\begin{aligned} F_{Y_i^*|X_i}(y^*|x) &= \Pr\left(Y_i^* < y^*|X_i = x\right) \\ &= \Phi\left(\frac{y^* - (\alpha_0 + \beta_0 x + \mu_0)}{\sigma_0}\right) \end{aligned}$$

where $\Phi(\cdot)$ denotes the CDF of a standard normal.

Question 2

The derivation of the discontinued cumulative distribution function $F_{Y_i|X_i}(y|x)$ relies on looking at the two cases of $y < 0$ and $y \geq 0$.
For $y < 0$, we know that $Y_i = Y_i^*$, and thus

$$\begin{aligned} \Pr(Y_i \leq y|x) &= \Pr(Y_i^* \leq y|x) \\ &= \Pr\left(\frac{Y_i^* - E[Y_i^*|x]}{\sigma_0} < \frac{y - E[Y_i^*|x]}{\sigma_0} \middle| x\right) \\ &= \Phi\left(\frac{y - (\alpha_0 + \beta_0 x + \mu_0)}{\sigma_0}\right) \end{aligned}$$

Notice, that we have used again $\frac{y^* - E[y^*|x]}{\sigma_0} \overset{d}{\sim} N(0,1)$.
For $y \in [0, \infty)$, we know that $P(y > 0|x) = 0$. Hence we have that $P(y \leq 0|x) = 1$.

Therefore, the CDF $F_{Y_i|X_i}$ is given by:

$$F_{Y_i|X_i}(y|x) = \begin{cases} \Phi\left(\frac{y-(\alpha_0+\beta_0 x+\mu_0)}{\sigma_0}\right), & y < 0, \\ 1, & y \geq 0. \end{cases} \tag{1}$$

Question 3

For the model to be identifiable we cannot have that

$$f(y|x;\alpha_0,\beta_0,\mu_0,\sigma_0) = f(y|x;\alpha_1,\beta_1,\mu_1,\sigma_1) \quad \forall(y,x)$$

where $\{\alpha_0,\beta_0,\mu_0,\sigma_0\} \neq \{\alpha_1,\beta_1,\mu_1,\sigma_1\}$. In words, we need that the true density of y can be described uniquely by the true parameters. However, in our case consider the set of true parameters $\boldsymbol{\theta_0} = \{\alpha_0,\beta_0,\mu_0,\sigma_0\}$ and the set of alternative parameters $\boldsymbol{\theta_1} \in \{\alpha_0+c,\beta_0,\mu_0-c,\sigma_0\}$, where $c \in \mathbb{R}$. From equation (1) it is obvious that we have

$$F_{Y_i|X_i}(y|x;\boldsymbol{\theta_0}) = F_{Y_i|X_i}(y|x;\boldsymbol{\theta_1}), \quad \forall(y,x)$$

which violates identifiability of the model. Intuitively, we cannot distinguish the effect of α_0 and μ_0, why we cannot find a unique pair that describes the density of Y_i.

Question 4

The conditional density of y_i can be described by the function,

$$f_{Y_i|X_i}(y \mid x;\boldsymbol{\theta}) = \begin{cases} 0, & y > 0 \\ P(y=0 \mid x), & y = 0, \\ f_{y|x,y<0}(y \mid x;\boldsymbol{\theta}), & y < 0, \end{cases}$$

First, we derive the conditional probability, using that $\alpha = 0$

$$\begin{aligned} \Pr(y=0|x) &= \Pr(\beta x + U_i > 0|x) \\ &= \Pr(U_i > -\beta x|x) \\ &= 1 - \Pr(U_i \leq -\beta x|x) \\ &= 1 - \Pr\left((U_i-\mu)/\sigma \leq -(\beta x+\mu)/\sigma|x\right) \\ &= 1 - \Phi\left(-(\beta x+\mu)/\sigma\right) \\ &= \Phi\left((\beta x+\mu)/\sigma\right), \end{aligned}$$

using that it is symmetric in the last step.

Second, the density for y, given that $y < 0$, is equivalent to the density of y^*. Thus, we can rewrite the density to a standard normal as follows,

$$\begin{aligned}
f_{y|x,y<0}(y \mid x; \boldsymbol{\theta}) &= f_{y_i^*|x}(y|x; \boldsymbol{\theta}) \\
&= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) \times \exp\left(-\frac{(y-(\beta x+\mu))^2}{2\sigma^2}\right) \\
&= \frac{1}{\sigma}\left(\frac{1}{\sqrt{2\pi}}\right) \times \exp\left(-\frac{(y-(\beta x+\mu))^2}{2\sigma^2}\right) \\
&= \frac{1}{\sigma}\phi\left(\frac{y-(\beta x+\mu)}{\sigma}\right),
\end{aligned}$$

where $\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)$, is the standard normal PDF.

Thus, we can write the model implied density:

$$f_{Y_i|X_i}(y|x; \boldsymbol{\theta}) = \left[\frac{1}{\sigma}\phi\left(\frac{y-(\beta x+\mu)}{\sigma}\right)\right]^{\mathbf{1}(y<0)} \left[\Phi\left(\frac{\beta x+\mu}{\sigma}\right)\right]^{\mathbf{1}(y=0)} \tag{2}$$

Question 5

Flipping the outcome yields $\tilde{Y}_i = -Y_i = -\min\{0, Y_i^*\} = \max\{0, -Y_i^*\}$. Thus, we can restate $\tilde{Y}_i$ as a tobit model, $\tilde{Y}_i = \max\{0, \tilde{Y}_i^*\}$, where $\tilde{Y}_i^* = -Y_i^*$. Let $\tilde{\beta} = -\beta_0$, $\tilde{\mu} = -\mu_0$ and $\tilde{\sigma} = \sigma_0$. Thus, the latent variable of the Tobit can be formulated $\tilde{Y}_i^* = \tilde{\beta}X_i + \tilde{U}_i$, where $\tilde{U}_i|X_i \sim N(\tilde{\mu}, \tilde{\sigma}^2)$. Inserting the Tobit parameters, using symmetry, and that $\tilde{y} = -y$, we derive the density from (2) as

$$\begin{aligned}
f_{\tilde{Y}_i|X_i}(\tilde{y}|x; \tilde{\boldsymbol{\theta}}) &= \left[\frac{1}{\tilde{\sigma}}\phi\left(\frac{-\tilde{y}+\tilde{\beta}x+\tilde{\mu}}{\tilde{\sigma}}\right)\right]^{\mathbf{1}(\tilde{y}>0)} \left[\Phi\left(\frac{-\tilde{\beta}x-\tilde{\mu}}{\tilde{\sigma}}\right)\right]^{\mathbf{1}(\tilde{y}=0)} \\
&= \left[\frac{1}{\tilde{\sigma}}\phi\left(\frac{\tilde{y}-(\tilde{\beta}x+\tilde{\mu})}{\tilde{\sigma}}\right)\right]^{\mathbf{1}(\tilde{y}>0)} \left[1-\Phi\left(\frac{\tilde{\beta}x+\tilde{\mu}}{\tilde{\sigma}}\right)\right]^{\mathbf{1}(\tilde{y}=0)},
\end{aligned}$$

which aligns with the associated conditional distribution in the Tobit model. Hence, the tiboT-model is exactly equavalent to the usual Tobit model stated in terms of $\tilde{Y}_i$, where the sign of β_0 and μ_0 is flipped.

Question 6

The usual maximum likelihood theory applies to Tobit (and therefore tiboT). For that reason, we can based on the density function in equation 2 write up the likelihood contribution of observation i as:

$$\ell_i(\theta) = \mathbf{1}\{y_i < 0\} \log \left[\frac{1}{\sigma} \phi \left(\frac{y_i - \mathbf{x}_i \boldsymbol{\beta} - \mu}{\sigma} \right) \right] + \mathbf{1}\{y_i = 0\} \log \left[\Phi \left(\frac{\mathbf{x}_i \boldsymbol{\beta} + \mu}{\sigma} \right) \right] \quad (3)$$

The estimation is not performed using the reparameterization described on slide 26 of lecture 17. Thus, the log-likelihood function is not globally concave, meaning that multiple local maxima may exist. Instead, the estimator is defined as the maximizer of the sum of the log-likelihood contributions.

In practice, we minimize the negative sum by employing the *nelder-mead* optimization alogrithm. The starting point for the optimization is a regular OLS of y on x where x besides regressor in dataset includes an intercept. The intercept is then interpreted as the starting value for μ, the slope as β and the standard error as σ. To avoid the aforementioned risk of ending in a local minima, we try various different starting values. None of which changes the argument that minimizes the negative sum of likelihood contribution. The reported standard errors are derived based on the outer product of the gradients as

$$\mathcal{I}(\hat{\theta}) = \frac{1}{N} \sum_{i=1}^{N} g_i(\hat{\theta}) g_i(\hat{\theta})^\top,$$

where $g_i(\hat{\theta}) = \frac{\partial \ell_i(\theta)}{\partial \theta}$ is the numerical gradient. The standard error of the estimated parameters is the diagonal of the covariance matrix.

$$\mathrm{Var}(\hat{\theta}) = \frac{1}{N} \mathcal{I}(\hat{\theta})^{-1} \quad \text{and} \quad \mathrm{SE}(\hat{\theta}_j) = \sqrt{\mathrm{Var}(\hat{\theta}_j)}.$$

For the validity of the calculated standard errors using the outer product of scores, the model must be correctly specified, meaning the likelihood function aligns with the true data-generating process—this condition is satisfied in this case. The data must also be independent and identically distributed (i.i.d.), with a sufficiently large sample size to ensure asymptotic properties hold, both of which are reasonable assumptions here. Additionally, the expectation of the score function at the true parameters must be zero, and the unconditional information matrix equality must hold. The latter requires the true parameters to be identified and interior to the parameter space, and the log-likelihood contributions, $\ell_i(\theta)$, to be continuous and continuously differentiable within the parameter space.

Lastly, the t-value of θ_j is computed as:

$$t_j = \frac{\hat{\theta}_j}{\mathrm{SE}(\hat{\theta}_j)}.$$

Parameter	$\hat{\beta}$	$\hat{\mu}$	$\hat{\sigma}$
Coefficient	1.437	0.543	1.536
SE	0.074	0.081	0.056
t	19.431	6.672	27.345

Tabel 1: Estimation results *showing coefficients, standard errors (SE), and t-statistics for the parameters $\hat{\beta}$, $\hat{\mu}$, and $\hat{\sigma}$.*

The results from the estimation is presented in table 1. All parameter estimation are significant on a level of confidence of 5 percent.

Question 7

From question 4, we have that the probability of censoring is given by:

$$\Pr(Y_i = 0|X_i = x) = \Phi\left(\frac{\beta_0 x + \mu_0}{\sigma_0}\right).$$

To derive an expression for the partial effect of the regressor $X_i = x$, we take the partial derivative. Using the chain rule, we get the result

$$\begin{aligned}\frac{\partial \Pr(Y_i = 0|X_i = x)}{\partial x} &= \frac{\partial}{\partial x}\left[\Phi\left(\frac{\beta_0 x + \mu_0}{\sigma_0}\right)\right] \\ &= \frac{\beta_0}{\sigma_0}\phi\left(\frac{\beta_0 x + \mu_0}{\sigma_0}\right).\end{aligned}$$

The number of partial effects are equal to the number of observations, i.

Question 8

To formally test whether X_i has zero partial effect of the probability of censoring, we test whether $\beta_0 = 0$. We do so by stating the null hypothesis $H_0 : \beta_0 = 0$ against the alternative $H_A : \beta_0 \neq 0$. To derive a test statistic, we compute the t-test

$$t = \frac{\hat{\beta}}{\text{SE}(\hat{\beta})}.$$

Under the null-hypothesis t is asymptotically standard normal, why if $|t| > 1.96$ we reject the null-hypothesis on a 5 percent confidence level. In our case the t-test statistic is $19.431 > 1.96$. We reject our null-hypothesis and conclude that on a 5 percent level of confidence, X_i has some partial effect on the probability of censoring.

Question 9

By law of iterated expectations

$$
\begin{aligned}
E(Y_i|X_i = x) &= \Pr(Y_i < 0|X_i = x)E(Y_i|X_i = x, Y_i < 0) \\
&\quad + \Pr(Y_i \geq 0|X_i = x)E(Y_i|X_i = x, Y_i \geq 0) \\
&= \Pr(Y_i < 0|X_i = x)E(Y_i|X_i = x, Y_i < 0) + 0 \qquad (9.1) \\
&= \big(1 - \Pr(Y_i = 0|X_i = x)\big)\, E\left(Y_i|X_i = x, Y_i < 0\right) \qquad (9.2) \\
&= \left[1 - \Phi\left(\frac{\beta_0 x + \mu_0}{\sigma_0}\right)\right] E(Y_i|X_i = x, Y_i < 0)
\end{aligned}
$$

Using that $Y_i = 0$, when $Y_i \geq 0$ in eq. (9.1) and recognizing that the complementary probability of $P(Y_i < 0|X_i = x)$ is $1 - P(Y_i = 0|X_i = x)$ in eq. (9.2).

Deriving the conditional mean we have that:

$$
\begin{aligned}
E(Y_i|X_i = x, Y_i < 0) &= E(\beta_0 x + U_i|X_i = x, \beta_0 x + U_i < 0) \\
&= \beta_0 x + E(U_i|X_i = x, U_i < -\beta_0 x) \\
&= \beta_0 x + E(U_i|U_i < -\beta_0 x) \qquad (9.3) \\
&= \beta_0 x + \mu_0 - \sigma_0 \frac{\phi\left(\frac{-(\beta_0 x + \mu_0)}{\sigma_0}\right)}{\Phi\left(\frac{-(\beta_0 x + \mu_0)}{\sigma_0}\right)} \qquad (9.4)
\end{aligned}
$$

where 9.3 is by independence and 9.4 is using the property of a truncated normal distribution on U_i. Inserting this result yields:

$$
\begin{aligned}
E(Y_i|X_i = x) &= \left[1 - \Phi\left(\frac{\beta_0 x + \mu_0}{\sigma_0}\right)\right] \left[\beta_0 x + \mu_0 - \sigma_0 \frac{\phi\left(\frac{-(\beta_0 x + \mu_0)}{\sigma_0}\right)}{\Phi\left(\frac{-(\beta_0 x + \mu_0)}{\sigma_0}\right)}\right] \\
&= \left[1 - \Phi\left(\frac{\beta_0 x + \mu_0}{\sigma_0}\right)\right] [\beta_0 x + \mu_0] - \sigma_0 \left[\Phi\left(\frac{-(\beta_0 x + \mu_0)}{\sigma_0}\right) \frac{\phi\left(\frac{(\beta_0 x + \mu_0)}{\sigma_0}\right)}{\Phi\left(\frac{-(\beta_0 x + \mu_0)}{\sigma_0}\right)}\right] \qquad (9.5) \\
&= \left[1 - \Phi\left(\frac{\beta_0 x + \mu_0}{\sigma_0}\right)\right] [\beta_0 x + \mu_0] - \sigma_0 \phi\left(\frac{(\beta_0 x + \mu_0)}{\sigma_0}\right),
\end{aligned}
$$

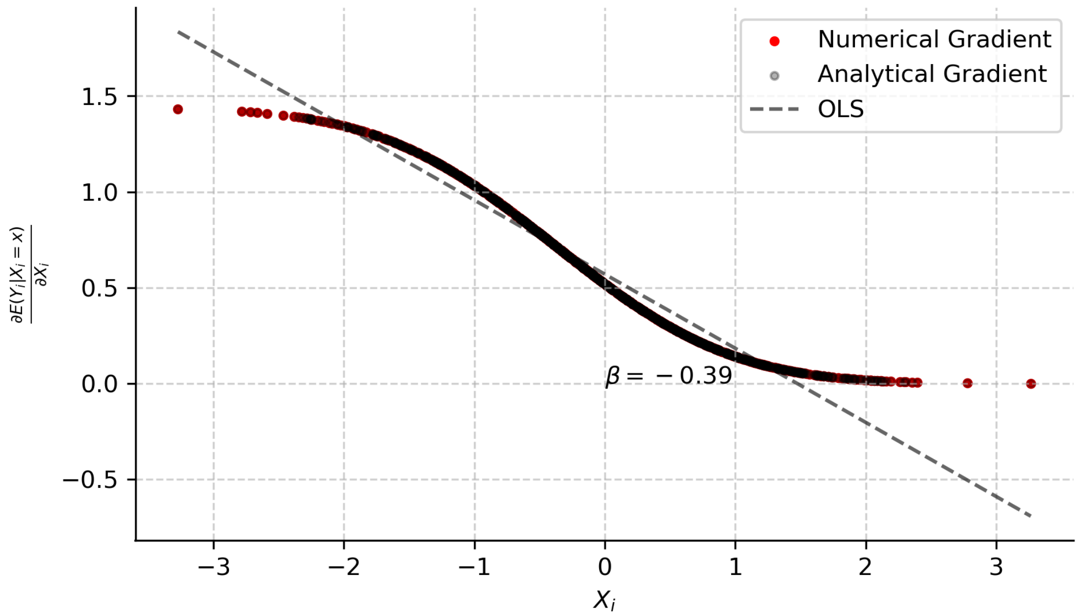
as required, with 9.5 following directly from the symmetry property.

Question 10

In the previous question, we derived the conditional mean of the observable Y_i given $X_i = x$ for the tiboT model. To test whether the partial effect of X_i on the conditional mean is constant, we compute the analytical partial derivative of the conditional mean with respect to X_i:

$$\text{PE}_i = \frac{\partial E(Y_i|X_i = x)}{\partial x} = \beta_0 \left[1 - \Phi\left(\frac{\beta_0 x + \mu_0}{\sigma_0}\right)\right],$$

This expression is postulated using the equivalent derivative of the Tobit model and the relationship between the models as shown in question 5. It thus serves as a hypothesized analytical form and to confirm the validity of this derivative, we also compute the numerical gradient, which aligns with the analytical result, see figure 1.



Figur 1: *Comparison of the numerical (red) and analytical (gray) gradients of* $\frac{\partial E(Y_i|X_i=x)}{\partial x}$*, with the OLS fit (dashed line) showing a significant negative slope. This demonstrates that* X_i *does not have a constant partial effect on the conditional mean.*

To test whether the partial effect is constant, we perform an OLS regression to examine whether the partial effect depends significantly on X_i. Specifically, we estimate the following regression model:

$$\widehat{\text{PE}}_i = \gamma_0 + \gamma_1 X_i + \varepsilon_i,$$

where $\widehat{\text{PE}}_i$ represents computed partial effect of X_i on Y_i, and ε_i is the residual. The coefficient γ_1 captures whether the partial effect varies systematically with X_i.

If the partial effect is constant, then $\gamma_1 = 0$, indicating no relationship between the partial effect and X_i. The null and alternative hypotheses are:

$$H_0 : \gamma_1 = 0, \quad H_1 : \gamma_1 \neq 0.$$

From the results shown in Figure 1, there is a strong negative correlation between X_i and the partial effect. Specifically, we estimate a coefficient of $\hat{\gamma}_1 = -0.39$, with a t-value of -82, indicating a highly significant relationship. This suggests that as X_i increases, the partial effect decreases, thus rejecting the null hypothesis H_0. Consequently, we conclude that X_i does not have a constant partial effect on the conditional mean.

Question 11

Although the model is misspecified we can still leverage the information about the conditional expectation. We propose an M-estimation procedure to estimate $\theta = \{\beta, \mu, \sigma\}$ based on the following minimization problem

$$\theta \in \arg \min_{\theta \in \mathbb{R}^3 | \sigma > 0} E[(Y_i - E[Y_i | X_i = x])^2]$$

and by sample analogue

$$\theta = \arg \min_{\theta \in \mathbb{R}^3 | \sigma > 0} \frac{1}{N} \sum_{i=1}^{N} (Y_i - m(X_i, \theta))^2, \quad \text{where}$$
$$m(X_i, \theta) = \left[1 - \Phi\left(\frac{\beta X_i + \mu}{\sigma}\right)\right] [\beta X_i + \mu] + \sigma \phi\left(\frac{(\beta X_i + \mu)}{\sigma}\right)$$

This is a Non-Linear-Least Squares (NLS) estimator. For consistency we rely on the theorem of (Newey and McFadden, 1994). If the minimization problem is uniquely solved by $\theta_0 = \{\beta_0, \mu_0, \sigma_0\}$, the objective function is convex, and the sample analogue converges in probability to the true moment of Y_i, then $\theta \xrightarrow{p} \theta_0 = \{\beta_0, \mu_0, \sigma_0\}$. The estimator is asymptotically normal if 1) θ_0 is an entirior solution, 2) $Y_i - m(X_i, \theta)$ is twice differentiable wrt θ, 3) The expected score $(\nabla_\theta m(X_i, \theta))$ equal 0, and 4) the expected hessian $(\nabla^2_\theta m(X_i, \theta))$ is positive definite.

Given that the conditional mean is well specified, the model satisfies $E(u_i | X_i = x) = 0$ with standard notation meaning that $u_i \equiv Y_i - m(X_i, \theta)$. Thus, we use the semi-robust estimator of the asymptotic variance of $\hat{\theta}$

$$\widetilde{\text{Avar}}(\widehat{\theta}) = \left(\sum_{i=1}^{N} \nabla_\theta \widehat{m}_i' \nabla_\theta \widehat{m}_i\right)^{-1} \left(\sum_{i=1}^{N} \widehat{u}_i^2 \nabla_\theta \widehat{m}_i' \nabla_\theta \widehat{m}_i\right) \left(\sum_{i=1}^{N} \nabla_\theta \widehat{m}_i' \nabla_\theta \widehat{m}_i\right)^{-1}.$$

and the standard errors are obtained as the square root of the diagonal elements of $\widetilde{\text{Avar}}(\hat{\theta})$.