Advanced Microeconometrics - Project 2

Group 21
Lukas Malte Kemeter (xtj271),
Felix Heuer (pfv999),
Malte Jacob Rattenborg (rnp562)

December 17, 2020

Note: The text for Part 2 amounts to five normal pages (Font size = 12p, 1.5 line spacing, and margins of 2.5 cm) according to the restrictions. The output amounts to two pages. For sake of readability, we include the tables and figures into the text.

Problem 1

(1):
The probability of y_i^* being censored corresponds to:

$$Pr(y_i^* \leq 0|x_i) = 1 - Pr(y_i^* > 0|x_i)$$

It states the probability, that the latent variable y^* is smaller or equal to zero and thus is be observed at the true value of the latent variable. If the latent variable is smaller or equal to zero, the obserserved variable equals 0. Thus the probability of being censored can be derived by the share or probability of observation being equal zero. Additionally, as all observed outcomes are greater or equal zero, it follows that:

$$Pr(y_i^* \leq 0|x_i) = Pr(y_i = 0|x_i) = 1 - Pr(y_i > 0|x_i)$$

Vice-verse, the probability of Y^* being uncensored can be expressed by:

$$Pr(y_i^* > 0|x_i) = Pr(y_i > 0|x_i) = 1 - Pr(y_i = 0|x_i)$$

(2):
The conditional cumulative distribution function (CDF) is derived based on the probabilities

from 1.1 and is calculated for different values of y_i as follows:

$$F_{Y|X}(y_i|x_i) = \begin{cases} F_1 \; if \; y_i < 0 \\ F_2 \; if \; y_i = 0 \\ F_3 \; if \; y_i > 0 \end{cases}$$

At first, as motivated in 1.1., all observed outcomes are weakly greater than zero and thus the cumulative density of y_i is 0 for all $y_i < 0$. Additionally, as $y_i = 0$ includes both censored latent variable and latent variables actually being zero, the cumulative density corresponds to some fixed value:

$$\begin{aligned} F_{Y|X}(0|x_i) &= Pr(y = 0|x_i) \\ &= Pr(y^* \leq 0|x_i) \\ &= Pr(\beta_0 x_i + \sigma_0 \varepsilon \leq 0|x_i) \\ &= Pr(\varepsilon \leq -\frac{\beta_0 x_i}{\sigma_0}|x_i) \\ &= G(-\frac{\beta_0 x_i}{\sigma_0}) \end{aligned}$$

Now we estimate the CDF for values of y_i above 0. We use the fact, that $Pr(y_i^* > 0|X = x_i) = Pr(y_i > 0|X = x_i, y_i > 0)$:

$$\begin{aligned} F_{Y|X}(y_i|x_i, y_i > 0) &= Pr(y \leq y_i|x, y_i > 0) \\ &= Pr(y^* \leq y_i|x_i) \\ &= Pr(\beta_0 x_i + \sigma_0 \varepsilon \leq y_i|x_i) \\ &= Pr(\varepsilon \leq \frac{y_i - \beta_0 x_i}{\sigma_0}|x_i) \\ &= G(-\frac{y_i - \beta_0 x_i}{\sigma_0}) \end{aligned}$$

This gives us the combined function for the CDF:

$$F_{Y|X}(y_i|x_i) = \begin{cases} G(-\frac{\beta_0 x_i}{\sigma_0}) \;\; if \; y_i = 0 \\ G(\frac{y_i - \beta_0 x_i}{\sigma_0}) \;\; if \; y_i > 0 \end{cases}$$

Combining the partially defined function above we receive the following equation for the cumulative density function:

$$F_{Y|X}(y_i|x_i) = \mathbb{1}\{y_i = 0\}G(-\frac{\beta_0 x_i}{\sigma_0}) + \mathbb{1}\{y_i > 0\}G(\frac{y_i - \beta_0 x_i}{\sigma_0})$$

The resulting CDF is partially defined for values of y being equal to zero or larger than zero. It becomes clear that the CDF for y equal to 0 does not depend on y. This shows that the CDF is a fixed value, as y does not change in this area. For values above 0, the CDF is increasing in the difference between y and the OLS predicted value of y, which is intuitive.

(3):

The likelihood function depends on the individual likelihood contributions which can be derived for the following cases depending on y_i:

$$f(y_i|x_i) = \begin{cases} f_1 \;\; if \;\; y_i < 0 \\ f_2 \;\; if \;\; y_i = 0 \\ f_3 \;\; if \;\; y_i > 0 \end{cases}$$

We examine the density function for each of these three cases individually.

Case 1 ($y_i < 0$)**:** Will never happen, such that $f_1 = 0 \, \forall \, x_i < 0$, as there are no observed outcomes.

Case 2 ($y_i = 0$)**:** This case occurs when the latent variable is equal or smaller than zero $y_i^* \leq 0$. Thus the density function if $y_i = 0$ equals the cumulative density function of the latent variable for $y_i^* \leq 0$:

$$\begin{aligned} f(y_i = 0|x_i) &= Pr(y_i = 0|x_i) \\ &= Pr({y^*}_i \leq 0|x_i) \\ &= Pr(\beta_0 x_i + \sigma_0 \varepsilon_i \leq 0|x_i) \\ &= Pr(\varepsilon_i \leq -\frac{\beta_0 x_i}{\sigma_0}|x_i) \\ &= G(-\frac{\beta_0 x_i}{\sigma_0}) \end{aligned}$$

Case 3 ($y_i > 0$)**:** Happens in case of $y_i > 0$, $y_i = {y^*}_i$ and thus $Pr(y \leq y_i|x_i) = Pr(y^* \leq y_i|x_i)$. In the following we use that the probability density function for the observed variable for $y_i > 0$, which equals to the probability density function of the latent variable.

$$f(y_i|x_i) = f({y^*}_i|x_i) \;\; \forall \, y_i > 0$$

To derive the density function of y_i for $y_i > 0$ we start by estimating the cumulative density function for $y_i > 0$. This can be derived as follows:

$$\begin{aligned}
F_{Y|X}(y_i|x_i, y_i > 0) &= Pr(y^* \leq y_i|x_i) \\
&= Pr(\beta_0 x_i + \sigma_0 \varepsilon \leq y_i|x_i) \\
&= Pr(\varepsilon \leq \frac{y_i - \beta_0 x_i}{\sigma_0}|x_i) \\
&= G(\frac{y_i - \beta_0 x_i}{\sigma_0})
\end{aligned}$$

In order to derive the PDF, we take the first derivative of the CDF w.r.t. y_i :

$$\begin{aligned}
f(y_i|x_i) &= \frac{\partial F(y_i|x_i)}{y_i} \\
&= g(\frac{y_i - \beta_0 x_i}{\sigma_0})\frac{1}{\sigma_0}
\end{aligned}$$

Together, the density can be partially described as follows:

$$f(y_i|x_i) = \begin{cases} G(-\frac{\beta_0 x_i}{\sigma_0}) \; if \; y_i = 0 \\ \frac{1}{\sigma_0} g(\frac{y_i - \beta_0 x_i}{\sigma_0}) \; if \; y_i > 0 \end{cases}$$

The likelihood contribution function is then given by:

$$f(y_i|x_i) = \mathbb{1}\{y_i = 0\}G(-\frac{\beta_0 x_i}{\sigma_0}) + \mathbb{1}\{y_i > 0\}\frac{1}{\sigma_0}g(\frac{y_i - \beta_0 x_i}{\sigma_0})$$

To derive the maximum-likelihood estimator, we define the likelihood function as:

$$l = \prod_{i=1}^{N} f(y_i|x_i)$$

The Log-Likelihood function is then given by :

$$\begin{aligned}
L &= \sum_{i=1}^{N} log f(y_i|x_i) \\
&= \sum_{i=1}^{N} [\mathbb{1}\{y_i = 0\} log G(-\frac{\beta_0 x_i}{\sigma_0}) + \mathbb{1}\{y_i > 0\} log \frac{1}{\sigma_0} g(\frac{y_i - \beta_0 x_i}{\sigma_0})]
\end{aligned}$$

The maximum-likelihood (ML) estimator is the parameter θ_0 that maximizes the Log-likelihood function. The estimator is equivalent when maximizing the log-likelihood or the likelihood function. In the following we use the log-likelihood function and the resulting ML estimator:

$$\Theta_0: \ arg\max_{\Theta_0} \sum_{i=1}^{N} [\mathbb{1}\{y_i = 0\} logG(-\frac{\beta_0 x_i}{\sigma_0}) + \mathbb{1}\{y_i > 0\} log\frac{1}{\sigma_0} g(\frac{y_i - \beta_0 x_i}{\sigma_0})]$$

(4):

In order to derive $E[Y|X = x]$, we split the expectations in parts and use the previously derived probability of censoring. Additionally, we use the fact that $E[Y|X,Y \leq 0] = E[Y|X,Y = 0] = 0$, such that the expected value for y given $y = 0$ equals zero:

$$\begin{aligned} E[Y|X = x] &= Pr(y \leq 0|X) * E[Y|X,Y \leq 0] + Pr(y > 0|X) * E[Y|X,Y > 0] \\ &= Pr(y > 0|X) * E[Y|X,Y > 0] \end{aligned}$$

Furthermore, we use the expression for Pr(y > 0 | X), which was derived above:

$$Pr(y > 0|X) = 1 - G(-\frac{\beta_0 x_i}{\sigma_0})$$

We also use E[Y|X, Y>0] which equals to the expected value of the latent variable conditional on Y>0.

$$\begin{aligned} E[Y|X,Y > 0] &= E[Y^*|X,Y^* > 0] \\ &= E[\beta_0 x_i + \sigma_0 \varepsilon | x_i, Y^* > 0] \\ &= E[\beta_0 x_i + \sigma_0 \varepsilon | x_i, \beta_0 x_i + \sigma_0 > 0] \\ &= (\beta_0 x_i + \sigma_0) E[\varepsilon | x_i, \varepsilon > -\frac{\beta_0 x_i}{\sigma_0}] \end{aligned}$$

By applying the rule for truncated expectations: $E[A|A > k] = \int_k^\infty a f_{A|A>k}(a)\, da$ on the conditional expected value of the error term we can use that:

$$E[\varepsilon|x, \varepsilon > -\frac{\beta_0 x_i}{\sigma_0}] = \int_{-\frac{\beta_0 x}{\sigma_0}}^{\sigma_0} t f_{\varepsilon|\varepsilon > -\frac{\beta_0 x}{\sigma_0}}(t)\, dt$$

Furthermore, we use the rule for truncated densities that: $f_{\varepsilon|\varepsilon > -\frac{\beta_0 x}{\sigma_0}} = \frac{g(t)}{1 - G(-\frac{\beta_0 x_i}{\sigma_0})}$. By plugging this term and the expression for $E[\varepsilon|x, \varepsilon > -\frac{\beta_0 x_i}{\sigma_0}]$ into $E[Y^*|X,Y^* > 0]$, we derive the required expression:

$$
\begin{aligned}
E[Y|X=x] &= Pr(y>0|X)E[Y|X,Y>0] \\
&= \left[1-G_0(-\frac{\beta_0 x_i}{\sigma_0})\right]\left[(\beta_0 x_i+\sigma_0)\int_{-\frac{\beta_0 x_i}{\sigma_0}}^{\infty} t\frac{g(t)}{1-G(-\frac{\beta_0 x_i}{\sigma_0})}\,dt\right] \\
&= \beta_0 x_i\left[1-G_0(-\frac{\beta_0 x_i}{\sigma_0})\right]+[1-G_0(-\frac{\beta_0 x_i}{\sigma_0})]\sigma_0\int_{-\frac{\beta_0 x_i}{\sigma_0}}^{\infty} t\frac{g(t)}{1-G(-\frac{\beta_0 x_i}{\sigma_0})}\,dt \\
&= \beta_0 x_i\left[1-G_0(-\frac{\beta_0 x_i}{\sigma_0})\right]+\frac{1-G_0(-\frac{\beta_0 x_i}{\sigma_0})}{1-G_0(-\frac{\beta_0 x_i}{\sigma_0})}\sigma_0\int_{-\frac{\beta_0 x_i}{\sigma_0}}^{\infty} t g(t)\,dt \\
&= \beta_0 x_i\left[1-G_0(-\frac{\beta_0 x_i}{\sigma_0})\right]+\sigma_0\int_{-\frac{\beta_0 x_i}{\sigma_0}}^{\infty} t g(t)\,dt
\end{aligned}
$$

(5):

The expected value of Y conditional on X was shown to be:

$$E[Y|X=x]=\beta_0 x_i[1-G_0(-\frac{\beta_0 x_i}{\sigma_0})]+\sigma_0\int_{-\frac{\beta_0 x}{\sigma_0}}^{\sigma_0} t\ g(t)\,dt$$

We can derive the marginal effect (ME) of x on y by using the Leibniz rule for taking the derivative of an integral. The marginal effect is given by the derivative of the expected value of y given x with respect to x:

$$
\begin{aligned}
ME(x) &= \frac{\partial}{\partial x}E[Y|X=x] \\
&= \beta_0\left[1-G_0(-\frac{\beta_0 x_i}{\sigma_0})\right]+\beta_0 x_i\left[g_0(-\frac{\beta_0 x_i}{\sigma_0})\frac{\beta_0}{\sigma_0}\right]+\sigma_0\left[-\frac{\beta_0 x_i}{\sigma_0}\frac{\beta_0}{\sigma_0}g_0(-\frac{\beta_0 x_i}{\sigma_0})\right]+\int_{-\frac{\beta_0 x_i}{\sigma_0}}^{\infty} 0 \\
&= \beta_0[1-G_0(-\frac{\beta_0 x_i}{\sigma_0})]+\beta_0 x_i[g_0(-\frac{\beta_0 x_i}{\sigma_0})\frac{\beta_0}{\sigma_0}]+\beta_0[-\frac{\beta_0 x_i}{\sigma_0}g_0(-\frac{\beta_0 x_i}{\sigma_0})] \\
&= \beta_0[1-G_0(-\frac{\beta_0 x_i}{\sigma_0})+g_0(-\frac{\beta_0 x_i}{\sigma_0})\frac{\beta_0 x_i}{\sigma_0}-\frac{\beta_0 x_i}{\sigma_0}g_0(-\frac{\beta_0 x_i}{\sigma_0})] \\
&= \beta_0[1-G_0(-\frac{\beta_0 x_i}{\sigma_0})]
\end{aligned}
$$

To examine how ME(x) depends on x can be seen by taking the FOC w.r.t. x:

$$
\begin{aligned}
&\frac{\partial}{\partial x}\beta_0[1-G_0(-\frac{\beta_0 x_i}{\sigma_0})] \\
&= \beta_0[g(-\frac{\beta_0 x_i}{\sigma_0})\frac{\beta_0}{\sigma_0}]>0
\end{aligned}
$$

As the PDS is defined to be larger than 0 (g(t)>0), this term is positive and we see that the ME(x) increases with x. Recall, that the ME(x) can only be interpreted at specific values of x. It thus reflect a non-linear relationship. Accordingly, x increases the ME as the impact of censored observations is reduced for higher values of x such that under non-linearity the effect becomes larger.

(6):

The claim can be evaluated by comparing the marginal effects of the latent variable and the censored ME calculated in (5). The latent outcome function shows that the marginal effect of x on the expected value of Y is β_0. The estimated marginal effect of x on y for the censored variable is: $\beta_0[1-G_0(-\frac{\beta_0 x}{\sigma_0})]$. The cdf function $G_0(t)$ is defined between 0 and 1. This means that the estimated ME is weakly smaller than β_0. The censoring probability was given by: $Pr(y=0|x_i)$. As shown in (1.2) we can rewrite this as: $G_0(-\frac{\beta_0 x}{\sigma_0})$. This means that the censoring probability reduces the calculated effect of ME(x). The reason for that is that with a higher censoring probability more latent variables are observed as zeros and are thus observed larger than they actually are. Because for all censored observations with $y_i = 0$ the marginal effect of x on y is zero by construction, a larger share of censored variables leads to a lower estimated marginal effect.

(7):

In order to estimate the marginal effect of x on y, and find a consistent estimator $\hat{ME}(x)$, we use the given ME(x) from (5). As a consistent estimator for the marginal effect ME(x) we suggest to use the Tobit model, as the observed variable is censored at 0. The tobit model assumes a normal distribution for the error term $\varepsilon \sim N(0,\sigma^2)$. To get the standard normal distribution we use the transformation that $\frac{\varepsilon}{\sigma} \sim N(0,1)$ This would give us the following *plug-in estimator*, using the consistency of $\hat{\theta}$ derived by MLE:

$$\begin{aligned}\hat{ME}(x) &= \hat{\beta}\left[1-\Phi(-\frac{\hat{\beta}x}{\hat{\sigma}_0\sigma})\right]\\ &= \hat{\beta}\left[\Phi(\frac{\hat{\beta}x}{\hat{\sigma}_0\sigma})\right]\end{aligned}$$

with Φ= CDF of the standard normal distribution.

For any m-estimator to yield consistent estimates, we require Theorem 5.1, which is based on identification. We require the estimator to be the unique minimizer to be identified. For a given value of x_0, we can test identification of $ME(x_0)$ by examining whether different values of the coefficients yield the same results. The tobit model would provide estimates for $\hat{\beta}_0$ and the combined term $\hat{\sigma}_0\sigma$. It would be possible to construct different values of $\hat{\sigma}_0$ and σ, which yield the same coefficient. This means the different parameters $\hat{\sigma}_0$ and σ are not identified while the combined term $\hat{\sigma}_0\sigma$ is identified. As β_0 is also the unique minimizer of the criterion function and is therefore identified and θ_0 is a consistent estimate, we conclude that $\hat{ME}(x)$ is consistent given the correct distribution assumption G(). This is because these three conditions cover all stochastic aspects of ME(x) which need to be estimated, or accounted for, and thus imply consistent estimation.

(8):
We have shown, that in the Tobit model the marginal effects can be estimated according to:

$$\hat{ME}(x_0) = \hat{\beta}\left(\Phi(\frac{\hat{\beta}x_0}{\hat{\sigma}_0\sigma})\right) \equiv h(\hat{\theta})$$

Where $h(\hat{\theta})$ is a plug a *plug-in estimator* with $\hat{\theta} = (\hat{\beta} \quad \hat{\sigma}_0)'$ representing a staggered version of the MLE parameter and G being the CDF of the standard normal distribution
In the following, the *Delta-Method* is used to derive the asymptotic distribution of the marginal effects estimator represented by the function h. Hence we are analysing $\sqrt{n}(h(\hat{\theta}) - h(\theta_0))$ which according the Delta-Method can be rewritten as:

$$\sqrt{n}\left(h(\hat{\theta}) - h(\theta_0)\right) = \frac{h(\hat{\theta}) - h(\theta_0)}{\hat{\theta} - \theta_0} * \sqrt{n}(\hat{\theta} - \theta_0)$$

The first part represents the rate of change w.r.t to $\hat{\theta}$ relative to its true value θ_0. the second term captures the asymptotic deviation of the estimate from its true value.
We use the fact that $h(\hat{\theta})$ is a continuous and differentiable function, [1] and the characteristics of $\hat{\theta}$, namely consistency and convergence in distribution according to:

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} Z \sim N(0, V_0)$$

Where V_0 is a 2x2 variance-covariance matrix between $\hat{\beta}$ and $\hat{\sigma}$. Additionally, the consistency of $\hat{\theta}$, identification of $\hat{\beta}$ and $\hat{\sigma}$ and the correct distribution assumption also implies consisteny of the plug-in estimator ($h(\hat{\theta}) \to h(\theta_0)$). Combining these assumption, one can follow the Delta-Method and show that:

$$\frac{h(\hat{\theta}) - h(\theta_0)}{\hat{\theta} - \theta_0} * \sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} \nabla h(\theta_0) * Z$$

Where Z follows from the asymptotic attributes of $\hat{\theta}$, and $\nabla h(\theta_0)$ is the the gradient of h at θ_0 containing the partial derivatives of h w.r.t. β & σ:

[1] We know G(.), the given CDF, to be continuous

$$\nabla h(\theta_0) = \begin{pmatrix} \frac{\partial h}{\partial \hat{\beta}_0} \\ \frac{\partial h}{\partial \hat{\sigma}_0} \end{pmatrix}' = \begin{pmatrix} (\Phi(\frac{\beta_0 x_0}{\sigma_0 \sigma}) + \hat{\beta}_0 \phi(\frac{\beta_0 x_0}{\hat{\sigma}_0 \sigma}) \frac{x_0}{\sigma_0 \sigma} \\ -\hat{\beta}_0 \phi(\frac{\hat{\beta}_0 x_0}{\hat{\sigma}_0 \sigma}) \frac{\hat{\beta}_0 x_0}{(\hat{\sigma}_0 \sigma)^2} \end{pmatrix}'$$

Additionally, given that Z is normally distributed and the fact that the normal distribution is closed under affine transformations, we can infer $h(\hat{\theta})$:

$$\sqrt{n}(h(\hat{\theta}) - h(\theta_0)) \xrightarrow{d} Z \sim N(0, \Omega)$$

Furthermore, with $\nabla h(\theta_0)$ being a vector of deterministic scalars the variance Ω can be stated by $\Omega = \nabla h(\theta_0)' V_0 \nabla h(\theta_0)$.

Thus we conclude the asymptotic distribution of the ME(x) estimator:

$$\sqrt{n}(h(\hat{\theta}) - h(\theta_0)) \xrightarrow{d} N(0, \nabla h(\theta_0) V_0 \nabla h(\theta_0)')$$

, with $h(\hat{\theta}) = \hat{M}E(x)$
(9):
In order to construct a confidence interval for the ME(x) estimation, it is required to have consistently estimate the limit variance Ω as well the following conditions:

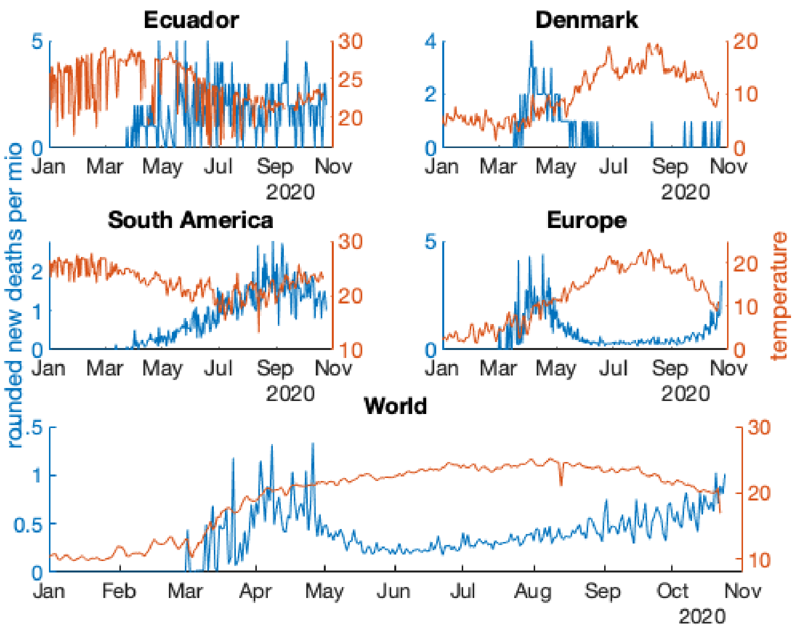
1. $\sqrt{n}(h(\hat{\theta}) - h(\theta_0)) \xrightarrow{d} N(0_{kx1}, \sum_0)$ for some $\sum_0 \in R_{KxK}$ symmetric positive definite,
2. h is continuously differentiable at θ_0, with derivatives $\nabla h(\theta_0)$ of full rank
3. $\hat{\sum}_0$ is a sequence of variance estimators consistent for $\sum_0$

The first requirement implies a consistent estimation of the marginal effects, and allows for the usage of the standard normal distribution for inference and the corresponding t-statistic through the distributional assumption. It is satisfied by the given specification of θ_0 being normal distributed. Additionally, variance-covariance matrices are positive semi-definite by definition. Assuming a strictly positive definite variance-covariance matrix is an additional restriction which cannot be confirmed nor denied at this point, but would have to be confirmed by the sample variance-covariance matrix. The second requirement ensures variance estimation, which is captured by the diagonal elements of the variance-covariance matrix. Continuous differentiability is satisfied by the underlying assumption of the given CDF $G(.)$, while a full rank of the gradient cannot be verified theoretically. At last, the used variance estimation can be assumed to be consistent due to the MLE estimation, given that the normal distribution assumption of G() is correct.

Problem 2

2.1 First, we are discuss the sample and covariate choices for analysing the effect of seasonal temperature changes on the number of COVID-19 deaths. Daily new COVID-19 deaths per million is our variable of interest to account for differences in population between countries. Following the idea by (Group 20, Project 2) we round the variable to the closest integer to fulfill the count data requirement of the Poisson model. The results in Table **??** indicate that the linear univariate relationship between temperature and new deaths per million is neither homogeneous across continents nor within continents. The effect is significant across all samples but only explains a faction of the underlying variation. The variation observed in the pooled regression can be divided into between and within variation. The later captures seasonal and daily weather changes, the primer different average temperatures between countries. The significant coefficient for the between estimator model indicates that different average temperatures between countries play an important role in predicting the number of deaths. The fixed effects (FE) estimator and the between estimator estimate different signs of the effect. This indicates that between variation reduces the within effect, which leads to a negative effect in the POLS model. Figure 1 shows the variation in temperature and new deaths for different countries and continents. We compare Europe and South America as both regions have very different trends of temperature and deaths due to COVID-19. There are notable differences in the within country variation in seasonal temperature. E.g. Ecuador shows a lot of day-to-day variation, while Denmark exhibits a steady seasonal weather trend over the year. Additionally, both countries have different yearly average temperate levels, representing notable between country variation in temperature. We account for country specific characteristics that are correlated with the average temperature by including a set of arguably exogenous covariates. It is necessary for these covariates not to be outcomes themselves, i.e. should not be affected by seasonal changes in temperature, but affect new deaths per million. To account for differences in the overall health system in a country we control for the number of hospital beds, female and male smokers, people with diabetes, the cardiovascular death rate and country-specific general life expectancy. Furthermore, as the death rate in a given country is related to the median age of the population and the population density, as people get infected faster in densely populated areas, we include these as covariates. We include GDP and the human development index to control for the general economic situation to account whether people can afford health care or staying at home from their job. We also account for political response, as stringent responses make people to stay at home. We assume stringency to be highly relevant in this model, even though it might be to some extend endogenous as politicians might react to temperature changes to reduce contact rates. Nevertheless, we assume that this can be neglected as number of cases should be the main driver. Hence, we include political stringency as a covariate to control for the exogenous variation in mobility, which is an endogenous variable. Finally, the used sample is restricted to the Northern hemisphere. We assume comparisons between the Northern and Southern hemi-

Figure 1: Variation between and within countries



sphere to be invalid as we would pick up on different country characteristics rather than within variation in temperature. Several countries which appear to have misreported deaths were also excluded. [2]

2.2

Tobit: To estimate the marginal effect (ME) of temperature on new deaths per million, we consider a Tobit model with an assumed standard normal distribution of the error term. The Tobit model applies due to a non-negative number of deaths. We are interested in the observed variable and model a theoretical latent variable to account for this corner solution property. Consistency requires homoskedasticity for the Tobit model. The used model is given by:

$$Y^* = \beta_0 X + \varepsilon, \;\; \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

$$Y = max[0, Y^*]$$

with Y^* being the latent variable, Y the deaths per million, X are covariates including temperature, and ε and the error term. The criterion function is given by:

$$q = -[\mathbb{1}\{y_i = 0\} log\Phi(-\frac{\beta_0 x_i}{\sigma}) + \mathbb{1}\{y_i > 0\} log\frac{1}{\sigma}\phi(\frac{y_i - \beta_0 x_i}{\sigma})]$$

The main features of interest for the Tobit model are the ME on the unconditional and conditional expected value. The marginal effects on the unconditional expectations are estimated according:

$$\hat{ME}(x) = \hat{\beta}\,[\Phi(\frac{\hat{\beta}x}{\hat{\sigma}})]$$

[2]We exclude Iran, Russia, Turkey and China from our sample. Iran: https://www.bbc.com/news/world-middle-east-53598965 Turkey: https://www.bloomberg.com/news/articles/2020-09-30/turkey-s-covid-wordplay-masks-extent-of-outbreak-lawmaker-says China: https://uk.reuters.com/article/instant-article/idUKKBN21Z052 Russia: https://www.cnbc.com/2020/05/20/russias-coronavirus-cases-top-300000.html

with Φ being the CDF of the standard normal distribution and $\hat{\beta}$ and $\hat{\sigma}$ the MLE estimates. The estimated ME on the conditional expected is given by:

$$\hat{ME}(x|y>0) = \hat{\beta}[1-\hat{\lambda}(x\hat{\beta}/\hat{\sigma}+\hat{\lambda})], \text{with } \lambda \text{ being the inverse Mills ratio}$$

Poisson: Furthermore, we estimate a Poisson regression model specified by Cameron and Trivedi (2005). The Poisson distribution is valid for non-negative integers and takes the following form:

$$Pr[Y=y] = \frac{e^{-\mu}\mu^y}{y!}$$

with $\mu = E[Y|X] = exp(x_i'\beta)$ being the conditional mean. To derive the ME of x on the expected value of Y, we take the partial derivative w.r.t to x.

$$\frac{\partial E[Y|x]}{\partial x} = \beta * exp(x\beta)$$

The ME estimator is then given by:

$$\hat{ME}(x)_{POISSON} = \hat{\beta} * exp(x\hat{\beta})$$

To estimate β and the ME we use the criterion function in Cameron and Trivedi (2005, p.118):

$$Q_N(\beta) = \frac{1}{N}\sum_{i=1}^{N}[-exp(x_i'\beta)+y_ix_i'\beta - lny_i!]$$

Next, we evaluate both models and their marginal effects at the mean value of temperature, as the ME are by definition non linear. Table 2 reports our estimated ME for the Tobit and the Poisson model in column 1 and 2, respectively. As the latent variable in Tobit model determines the censoring probability at 0, one can differentiate between ME on $E[Y|x]$ and $E[Y|x,y>0]$. For the Poisson model we only estimate the ME on the unconditional expected value. The Delta method is used to estimate standard errors for the ME and to conduct inference. To conduct the Delta method[3], we use the expression:

$$\sqrt{n}(ME(\hat{\beta}) - ME(\beta_0)) \xrightarrow{d} N(0, \nabla ME(\beta_0)V_0\nabla ME(\beta_0)')$$

where $\nabla h(\theta_0) = \frac{\partial ME(x)}{\partial \hat{\beta}}$ and V_0 as the variance-covariance matrix. Hence we obtain the analytical derivative of the ME's w.r.t the estimator $\hat{\beta}$. The analytical derivations are given in Table 1:

Table 1: Analytical Derivatives of ME's w.r.t β

	Tobit	Poisson
$\frac{\partial ME(x)}{\partial \hat{\beta}}$	$\left[\Phi(\frac{\hat{\beta}x}{\hat{\sigma}})\right] + \hat{\beta}\left[\phi(\frac{\hat{\beta}x}{\hat{\sigma}})\right]\frac{x}{\hat{\sigma}}$	$\hat{\beta}exp(x\hat{\beta})x + exp(x\hat{\beta})$
$\frac{\partial ME(x\|y>0)}{\partial \hat{\beta}}$	$(1-[2\hat{\beta}\frac{\hat{\lambda}x}{\sigma} + \hat{\beta}^2\frac{x}{\sigma}\hat{\lambda}'] - [\hat{\lambda}^2 + \hat{\beta}[2\hat{\lambda} - \hat{\lambda}^2])$	-

Note that $\hat{\lambda}' = [-\lambda(\frac{\hat{\beta}x}{\hat{\sigma}})[\frac{\hat{\beta}x}{\hat{\sigma}} + \lambda(\frac{\hat{\beta}x}{\hat{\sigma}})]]$

We can see that the unconditional ME of temperature is significantly positive for the Tobit model and negative for the Poisson model while the conditional ME is insignificantly different from 0 for the Tobit model. The difference in signs for the conditional ME suggests that the distributional assumptions of the two models make a strong difference for the estimated ME.

[3]See Part 1 Ex. 9 for more elaborate description.

Table 2: Marginal Effects

	Tobit	Poisson	Logistic	Extreme Value	Student t	CLAD
Temperature	0.044***	-0.005**	0.046***	0.055***	0.026***	-0.038
	(0.006)	(0.002)	(0.005)	(0.006)	(0.006)	(-)
$\frac{dE(y\|x,y>0)}{dx}$	0.010					
	(0.013)					
$\frac{dE(y\|x)}{dx}$	0.008***	-0.001***	0.008***	0.020***	0.005***	
	(0.001)	(0.0004)	(0.001)	(0.003)	(0.001)	
R^2	0.021	0.078	0.243	0.431	0.100	0.118
$Pr(y=0\|X)$	0.808	0.691	-	-	-	-
Flag	1	1	1	1	1	1

Standard errors in parenthesis and p<0.1*, p<0.05**, p<0.01***
Flag 1: 'Optimization terminated: the current x satisfies the termination criteria using OPTIONS.TolX of 1.0e-08 and F(X) satisfies the convergence criteria using OPTIONS.TolFun of 1.0e-08'
In case of the Poisson Model, the stated R^2 has to be interpreted with caution as this measure does not extend to Poisson regression.

2.3

A first impression of the underlying model fit can be observed in Figure 2, which displays the distribution of the observed realizations as well as the corresponding conditional model predictions. The predicted values are relatively good approximates of the observed outcomes. However, it appears that both models do not predict the number of deaths to a high extend. It appears that temperature shows only a weak prediction power for the number of COVID-19 deaths. A R^2 of 0.021 for the Tobit model and 0.078 for the Poisson model underline the low prediction power of the models. Furthermore, we compare the prediction power of the models in terms of the censoring probability $Pr(y=0|x)$. Censoring probabilities are estimated by:

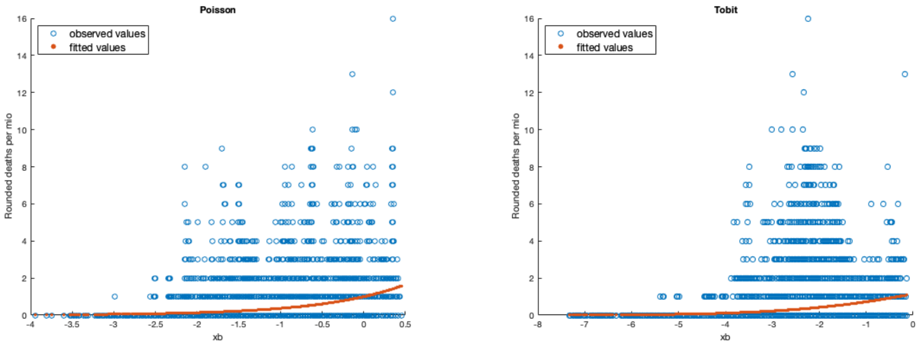
$$Tobit: Pr(y_i=0|x) = Pr(y_i^* \leq 0|x_i) = 1-\Phi(\frac{\beta_0 x_i}{\sigma})$$

$$Poisson: Pr(y_i=0|x) = \frac{e^{-\mu}\mu^0}{0!} = e^{-\mu}$$

With an estimated censoring probabilities of $Pr^{tobit}\big(y=0|X\big) = 0.808$ and $Pr^{poisson}\big(y=0|X\big) = 0.691$. The observed probability of $y=0$ in the data is 0.786. Hence, both models yield a fairly good prediction for the censoring probability. By taking into account observation equal zero differently, the Tobit model exhibits a better fit.

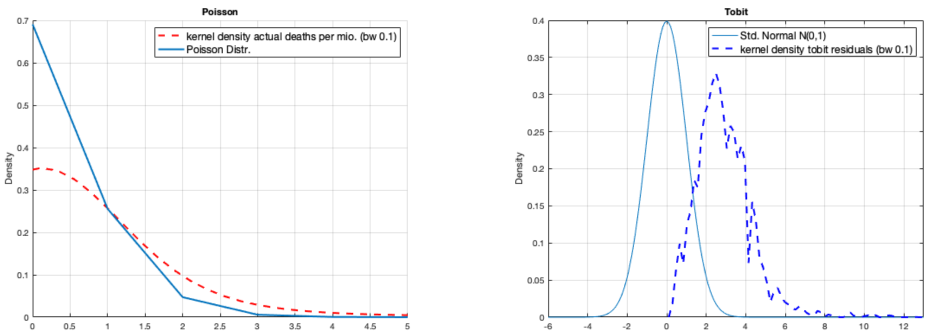
Figure 3 assess the validity of the underlying distributional assumptions. The left graph in Figure 3 compares the assumed Poisson distribution outcome variable with the observed density. This shows that the assumed density of the Poisson model is a good fit of the observed density. For values below 0, we do not have any observations due to censoring. For the Tobit model, we assume that the error terms follow a normal distribution with mean zero. The corresponding distribution of the Tobit residuals in Figure 3 appears not to be centered around zero and

Figure 2: Model Fit



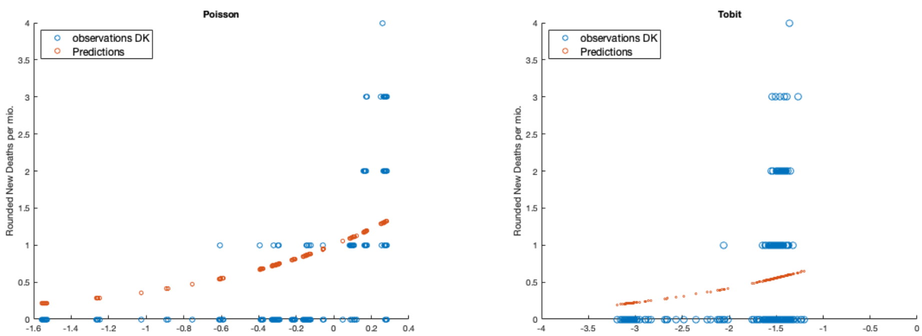
shifted to the right, which would violate the distribution assumption. If this results in a incorrectly specified criterion function, this questions consistent estimation of the Tobit model. This indicates that in general the underlying assumptions hold more closely for the Poisson model.

Figure 3: Distributions



Finally, Figure 4 shows the predictions of the Poisson and Tobit model relative to the observed values for Denmark respectively. The predictions are based on the full sample for the Northern hemisphere. Both models appear to predict the number of new deaths per million for higher temperature too low. The Poisson Model does a slightly better job in capturing the rapid increase in deaths for higher values of $x\beta$. The underestimation could potentially be explained by trend differences in the number of deaths in Denmark and the whole sample. We conclude superior predictions for Denmark for the Poisson model.

Figure 4: Model Fit Denmark



2.4

We run four alternative models to asses the robustness of the Tobit estimates w.r.t. the distribution of the error term.

CLAD: At first, the CLAD model accounts for heteroskedasticity and is given by:

$$Y^* = \beta_0 X + \varepsilon, \;\; Med(\varepsilon|x) = 0$$

$$Y = max[0, Y^*]$$

From Figure 2 it appears that the variation differs for larger values of x, so some of the observed variation might actually be heteroskedasticity. The CLAD defines the prediction function partially and therefore allows for heteroskedasticity. This yields the criterion function:

$$q = |Y - max(0, \beta_0 X)|$$

The CLAD model does not fulfill the assumption of twice differentiability, which is necessary to assume asymptotic normality of the estimator. Nevertheless, as illustrated in the lecture, the criterion function becomes smooth when $N \to \infty$. Hence, the function is twice differentiable for $N \to \infty$ and we can assume asymptotic normality of the estimator. Nevertheless, the results for the CLAD estimator do not report standard errors because numerical optimization is not possible for Theorem 5.3 under asymptotic normality. Based on the median distribution of the CLAD model the derivations of the ME according to the distribution assumed in the Tobit model is invalid. Therefore, compare the parameter coefficients between these models to infer whether the models differ strongly. The CLAD model hinges in the possibility to estimate marginal effects. To assess the robustness of the estimated ME of temperature on the unconditional expected value we run three alternative models. The comparison of the ME on the conditional expected value is not feasible in this case, as we took advantage of the special features of the normal distribution to derive the expression for the conditional expected value. We compare the unconditional MEs of these models. The first two alternative distributions follow suggestions from (Group 20, Project 2) who point out the advantages of using the logistic and the extreme value distribution. The logistic distribution has more mass on the tail of the distribution than the normal distribution. Hence, Figure 2 motivates the use of the logistic distribution due fatter tails of the density. Furthermore, we use the extreme value distribution, which is applicable to model the minimum of a variable and resembles a normal distribution closely. To assess the assumption of an error with an expected value of 0 it is reasonable to treat it as the allowed minimum of the error term. At last, the student-t distribution is close to the normal distribution, but has lower mass at the mean for N not going to infinity. Figure 3 motivates this distribution, as we see that the error terms have a lower density at the mean than predicted by the normal distribution.

The results of these robustness checks can be seen in Table 3, columns (3)-(6). For the logistic, extreme value and student t distribution we observe that the estimated MEs are close to the Tobit model and show the same sign. This indicates that the results of the Tobit model are robust to the assumed distribution. Nevertheless, the CLAD model shows a negative sign of the coefficient of interest. This suggests that the Tobit model might capture heteroskedasticity instead of the underlying relationship, which would lead to inconsistency of the Tobit model.

2.5

Table 4 shows the results of the Tobit model for different samples and the first and second half of the year in the northern hemisphere. In the following, we discuss the unconditional marginal effects. We see that the sign and magnitude of the effect of interest differs between Scandinavia and Southern Europe, and the Northern hemisphere. This indicates that there are pronounced heterogeneous effects between countries. Additionally, the relationship between temperature and COVID-19 deaths appears to be dependent on the underlying observation period. While for the first half, from January-May, the ME appears to be positive, in the second half until October the effect is insignificantly different from 0. For the winter months, the model predicts decreasing COVID-19 deaths due to falling temperatures in the northern hemisphere, ceteris paribus. But due to heterogeneous effects between countries would expect an increase in COVID-19 deaths due to decreasing temperatures in the winter, e.g. in Scandinavia or Southern Europe. The model predicts increasing deaths due to an increasing temperature next spring, as indicated by the significant positive ME in the 1^{st} half. Nevertheless, the captured effect might represent another underlying time trend. As cases increased exponentially in the spring months, when temperature increased, the estimated effect might capture missing information of governments about the pandemic. The insignificant ME in the summer months could represent the learning effect, e.g. improved treatment quality. Therefore, we do not assign an unambiguously causal interpretation to this estimated ME.

Table 3: Tobit Robustness checks

	Northern Hem.	North America	Scandinavia	Southern Europe	1st half [1]	2nd half [1]
Temperature	0.044***	-0.019	-0.146***	-0.067***	0.025***	-0.003
	(0.006)	(0.025)	(0.031)	(0.023)	(0.012)	(0.006)
dE(y|x,y>0)/dx	0.009	-0.004	-0.018	-0.014	0.003	-0.001
	(0.013)	(0.091)	(0.231)	(0.124)	(0.066)	(0.016)
dE(y|x)/dx	0.008***	-0.003	-0.005***	-0.012***	0.001***	-0.001
	(0.001)	(0.003)	(0.0001)	(0.001)	(0.0004)	(0.001)
Flag	1	1	1	1	1	1

Flag 1: 'Optimization terminated: the current x satisfies the termination criteria using OPTIONS.TolX of 1.0e-08 and F(X) satisfies the convergence criteria using OPTIONS.TolFun of 1.0e-08'
Table shows Tobit marginal effects.
[1] First half is January - May, second half covers observations from June - October and refer to the full sample (the northern hemisphere).

Bibliography

Cameron, A Colin and Pravin K Trivedi (2005). *Microeconometrics: methods and applications*. Cambridge university press.