Information about Individual Contributions

MUST be submitted along with the exam submission

According to Section 15 of the Examination Order, an individual grade or other assessment can only be given in a written group exam if the contribution of each student can be identified (individualisation). If the contribution of each group member is not specified, the exam submission will be rejected.

To assess a group assignment, it is therefore necessary to indicate who is responsible for which sections of your exam submission. This should be specified per section/chapter, not per line. Introduction, partial conclusions, conclusion, and appendices with calculations can be written collaboratively.

Please indicate the number of students in the group: **3**

Please provide information on section responsibility in exam number order:
(If the exam is not anonymous, please provide your name instead of exam number)

Exam number:	Responsible for the sections/chapters:
Konrad Juel Thide	Part 1: 1, 2.2, 4 Part 2: Q3, Q6, Q9
Asbjørn Fyhn	Part 1: 2, 2.3 Part 2: Q1, Q4, Q7, Q10
Theodor Justus Bock	Part 1: 2.1, 3 Part 2: Q2, Q5, Q8, Q11

Part 1: Assignment 1

Character count is around 12.000 without tables

12. januar 2025

1 Introduction

Firms are frequently assumed to achieve proportional increases in output by proportionally increasing inputs. This specific relationship between inputs and output is known as constant returns to scale. This paper empirically investigates the Cobb-Douglas production function, analyzing the influence of capital and labor in French manufacturing firms. We examine the input mix of manufacturing firms by employing fixed-effects and first-difference estimations. However, we are unable to develop a consistent estimator that allows us to formally assess the hypothesis of constant returns to scale. As if the estimator is consistent, we evaluate the hypothesis using first-difference estimation and reject it.

2 Econometric Theory

We formulate the relationship in the log-transformed Cobb-Douglass production function for firm i at time t as

$$y_{it} = \beta_K k_{it} + \beta_L l_{it} + c_i + u_{it}. \tag{1}$$

Here y_{it} denotes log of adjusted sales, k_{it} is log of adjusted capital stock and l_{it} is the log of employment. Hence, $\beta_K + \beta_L$ are input elasticities.

We employ a panel dataset of $N = 441$ French manufacturing firms observed at the odd years from 1968 to 1979. Given our panel data structure, we allow some time-invariant c_i to affect output that may vary across firms. As introduced, the objective of this analysis is to investigate the return to scale for capital and labor. In order to do so, we test whether $\beta_K + \beta_L = 1$, which would constitute constant return to scale.

In the following section, we introduce the estimators used for β_K and β_L. We present the required assumptions for the estimator(s) to be consistent and the implications for our estimator(s). Later, we formulate how these imposed assumptions can be formally tested. Finally, we present how we test whether there is evidence in the data for constant return to scale.

2.1 Estimation method

We seek to derive estimates of (β_K, β_L) by applying FE and FD estimation for equation (1). In the empirical analysis we outline our reasoning for the use of these estimators rather than POLS or RE estimation. The FE-estimator is given by,

$$\hat{\boldsymbol{\beta}}_{FE} = (\hat{\beta}_K, \hat{\beta}_L)'_{FE} = \boldsymbol{(\ddot{X}'\ddot{X})^{-1}\ddot{X}'\ddot{y}} \tag{2}$$

where $\ddot{\boldsymbol{X}}$ ($NT \times 2$) are the time-demeaned regressors (l_{it}, k_{it}) stacked as columns, and $\ddot{\boldsymbol{y}}$ with dimensions $NT \times 1$ is the time-demeaned dependent variable y_{it} stacked over t and i. The FD-estimator is

$$\hat{\boldsymbol{\beta}}_{FD} = (\hat{\beta}_K, \hat{\beta}_L)'_{FD} = \boldsymbol{(\Delta X'\Delta X)^{-1}\Delta X'\Delta y} \tag{3}$$

where $\boldsymbol{(\Delta X, \Delta y)}$ denotes (l_{it}, k_{it}) first-differenced and stacked as columns over t and i and y_{it} first-differenced and stacked over t and then i. Notice the first-differenced regressors imply that the dimensions of $\boldsymbol{\Delta X}$ is now $N(T-1) \times 2$ since we loose an observation when taking the first difference.

2.2 Consistency and asymptotic normality

For consistent estimators $\hat{\beta}_{FD}, \hat{\beta}_{FE} \xrightarrow{p} (\beta_K, \beta_L)$, we require that FE.1-2 and FD.1-2 hold. FE.1 imposes that $E[u_{it}|(k_{i1}, l_{i1}), (k_{i2}, l_{i2}), ..., (k_{iT}, l_{iT}), c_i] = 0, \forall t = 0, 1, .., T$, while FD.1 similarly requires $E[u_{it}|(k_{it}, l_{it}), c_i], \forall t = 0, 1, ..., T$. Hence, FE.1 and FD.1 are similar in nature and require exogeneity of the regressors, which we test formally. Although FE.1 is more strict than FD.1, we expect that FD.1 is also violated if FE.1 is rejected.

FE.2 requires that $rank(E[\boldsymbol{\ddot{X}_i'\ddot{X}_i}]) = 2$ while similarly for FD.2, $rank(E[\boldsymbol{\Delta X_i'\Delta X_i}]) = 2$, where $\boldsymbol{\ddot{X}_i}$ is $(T \times 2)$ and $\boldsymbol{\Delta X_i}$ is $((T-1) \times 2)$. These are merely necessary to identify $\hat{\beta}_{FE}$ and $\hat{\beta}_{FE}$ from equations (2) and (3). The following section will outline how we test the validity of these assumptions. Under FD.1 and FD.2 we have that

$$\sqrt{N}(\hat{\boldsymbol{\beta}}_{FD} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}\left(0, \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}\right),$$

where $\boldsymbol{A} = E\left(\boldsymbol{\Delta X}_i'\boldsymbol{\Delta X}_i\right)$, $\boldsymbol{B} = E\left(\boldsymbol{\Delta X}_i'\boldsymbol{e}_i\boldsymbol{e}_i'\boldsymbol{\Delta X}_i\right)$ and $\boldsymbol{e}_i = \boldsymbol{\Delta u_i}$ is the $((T-1) \times 1)$ error vector in FD-estimation. Similar asymptotic property holds for the FE-estimator under FE.1 and FE.2.

Finally, we have FE.3 which states that u_{it} are serially uncorrelated and have constant variance. This is similar for first difference, but here it its assumed for the first differenced errors, e_{it} where $e_{it} \equiv \Delta u_{it}$. These assumptions are important in terms of efficiency. The estimator for the robust variance matrix for the first difference regression model is:

$$\widehat{Avar(\hat{\boldsymbol{\beta}}_{\boldsymbol{FD}})} = (\boldsymbol{\Delta X'\Delta X})^{-1} \left(\sum_{i=1}^{N} \boldsymbol{\Delta X_i}' \hat{\boldsymbol{e}}_{\boldsymbol{i}} \hat{\boldsymbol{e}}_{\boldsymbol{i}}' \boldsymbol{\Delta X_i} \right) (\boldsymbol{\Delta X'\Delta X})^{-1} \tag{4}$$

where $\hat{\boldsymbol{e}}_{\boldsymbol{i}}$ is a $(T-1)\times 1$ vector with residuals of the regression model as defined above. Notice, it is equivalent for the FE regression model but with corresponding regressors and residuals.

2.3 Empirical tests

We test for strict exogeneity by conducting FE estimation on the two models, respectively

$$y_{it} = \beta_K k_{it} + \beta_L l_{it} + \delta_K k_{it+1} + c_i + u'_{it}$$
$$y_{it} = \beta_K k_{it} + \beta_L l_{it} + \delta_L l_{it+1} + c_i + u''_{it}$$

The null hypothesis of the test is that $\delta_K = 0$ or $\delta_L = 0$ against the two-sided alternative by a usual t-test. The test statictic is asymptotically standard normal under the null. If we reject the null hypothesis, we cannot assume strict exogeneity, which invalidates the FD- and FE-estimator.

We test for homoskedasticity by examining any potential autocorrelation in the error term. Specifically, we run the residual regression

$$\hat{u}^j_{it} = \rho^j \hat{u}^j_{it-1} + \epsilon^j_{it} \tag{5}$$

where $\hat{u}^j = \hat{u}$ for FE- and $\hat{u}^j = \hat{e}$ for FD-estimation. We compute standard t-tests with robust standard errors of the null hypothesis of $\rho^j = 0$, with the alternative that $\rho^j \neq 0$. The test statistics are asymptotically normal under the null.

Finally, we are interested in testing linear hypotheses. Formally, we test a null hypothesis of the form $\boldsymbol{R\beta} = \boldsymbol{r}$ where $\boldsymbol{R}$ is a $Q\times K$ matrix with Q being the number of restrictions and K number of regressors (i.e. 2 in our case) and $\boldsymbol{r}$ being a $Q\times 1$ matrix. The alternative hypothesis is $\boldsymbol{R\beta} \neq \boldsymbol{r}$. The Wald statistics for testing the above hypothesis is:

$$W = (\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})' \left[\boldsymbol{R} \widehat{Avar(\hat{\boldsymbol{\beta}})} \boldsymbol{R}' \right]^{-1} (\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})$$

where $Avar(\hat{\boldsymbol{\beta}})$ is the robust asymptotic variance of the estimated coefficients. The Wald statistics is asymptotically χ^2_Q-distributed with Q degrees of freedom under the null hypothesis. The following section will outline our empirical findings.

3 Empirical Findings

In terms of our choice of estimator we argue that both the POLS and RE estimation will be misleading in this framework. Specifically, if we believe that we have any time-constant effect within firms, i.e. $c_i \neq 0$ in eq. (1), the POLS will suffer from omitted variable bias. This could for instance be patents, if we believe that they are constant over time and affects firms output, which would lead to inconsistent estimates of our parameters. Furthermore, if e.g. firms with patents has higher levels of capital or labor, the RE assumption is violated, i.e. $E(c_i|k,l) \neq 0$, which would also lead to inconsistent estimates of our parameters.

Thus, we choose to estimate eq. (1) using FE and FD, as they allow for time-invariant productivity factors that may be influenced by the levels of capital and labor. We transform our data as stated in section 2.1, and estimate the coefficients by employing equation (2) and (3). The estimated coefficients are presented in Table 1 and 2. Both models are estimated using the estimator for the robust variance matrix from equation (4) as the residuals seem to suffer from autocorrelation in the residuals. The tests are conducted by running the auxiliary regression described by equation (5). For both models, we reject that $\rho^j = 0$ as the t-test statistics are 6.6 and -7.1 for the FE and FD models respectively, see Table 3.

We conduct test of the assumptions for strict exogeneity, as outlined in section 2.2. The test shows that the coefficients δ_K and δ_L are statistically significant on a 5% percent significance level. This is a major drawback for both models, as we reject consistency, and thus the estimates are not of interest. Thus, it is not possible to test for constant returns to scale. This is our primary finding. Rather than ending our assignment here, we continue as if the assumptions are not violated from section 2.2 to demonstrate the approach.

We argue that the FD estimator is preferred to the FE. Under FE.1-2 and FD.1-2, the choice between these estimators hinges on efficiency. More specifically, how we think the idiosyncratic errors, u_{it}, are distributed. In general, the FE estimator is proven more efficient when the u_{it} is serially uncorrelated, while the FD estimator is preferred when the error follows a random walk. Within the Cobb-Douglass model framework, the unobserved technology that is dependent on time, can be represented by u_{it}. Although we test and reject both assumptions, we argue that technology changes are more likely to follow a random walk rather than being serially uncorrelated. It seems implausible that the level of technology is uncorrelated with the previous level. Therefore, we believe that FD gives a better representation of the error term and is therefore our preferred model. Thus, if the estimator was in fact consistent, we would prefer the FD estimator.

The goal of this assignment is to test whether that product exhibit constant returns to scale. Constant return to scale would imply that the coefficients in the Cobb-Douglas would sum to 1. So in order to test the hypothesis of production exhibiting constant returns to scale, we formulate the null-hypothesis that $\beta_L + \beta_K = 1$, which corresponds to $\boldsymbol{R} = (1, 1)$ and $\boldsymbol{r} = (1)$, and the alternative hypothesis that $\beta_L + \beta_K \neq 1$. The test is conducted as explained in section 2.3. We get a Wald test statistics of 37.35 with a p-value of 0.0. This means that we on a 5% significance level can reject that the test statistics is χ^2_1-distributed and therefore that the $\beta_L + \beta_K = 1$. The economic interpretation is that we reject the hypothesis that production exhibits constant return to scale.

4 Conclusion and discussion

In this assignment, we explored the Cobb-Douglas production function using both fixed effects (FE) and first differences (FD) estimators. Both estimators failed the assumptions of strict exogeneity, which makes them inconsistent and thus we are unable to apply valid inference. This issue might stem from a relationship between the time-dependent total factor productivity (TFP) and the labour and capital stock. If smaller firms can more easily adapt new technology, or alternatively larger firms benefit from increasing know-how in production, the imposed relationship between k, l and u is violated. Additionally, if k_{t+1} is increasing in past profits and thereby TFP-shocks $u_{it}, ..., u_{i0}$, the assumption of strict exogeneity does not hold. We find such a dynamic relationship likely. It is however difficult to combat the issue of exogeneity since TFP shocks are unobservable in nature.

Despite this econometric challenge, we test the hypothesis of constant returns to scale. The empirical results reject this hypothesis, suggesting that production does not exhibit constant returns to scale. However, the reliability of our conclusion depends on the appropriateness of the models. Even though we prefer the FD-estimator in terms of efficiency, neither model fully captures complex factors like technological changes, introducing inconsistency.

First Difference

	Coefficient	*Std. err.*	*test statistic*	*P-value*
β_L	0.73	0.03	21.44	0.00
β_K	0.05	0.03	2.04	0.04
R^2	0.47			

Tabel 1: First difference estimation output. Table presents the estimated coefficients, standard error, t-test stastictics and the P-value (i.e. the probability of the estimated coefficient being insignificant

Fixed Effect

	Coefficient	*Std. err.*	*test statistic*	*P-value*
β_L	0.71	0.03	25.04	0.00
β_K	0.14	0.02	6.36	0.00
R^2	0.47			

Tabel 2: Fixed Effect estimation output. Table presents the estimated coefficients, standard error, t-test stastictics and the P-value (i.e. the probability of the estimated coefficient being insignificant

Hypothesis	**Test stastistic**	**P-value**	**Conclusion**
Strict Exogeneity (lemp)	$t = 2.8$	$P = 0.5\%$	Reject Null (Strict Exogeneity does not hold)
Strict Exogeneity (lcap)	$t = 5.3$	$P = 0\%$	Reject Null (Strict Exogeneity does not hold)
No Serial Autocorrelation (FE Model)	$t = 6.7$	$P = 0\%$	Reject Null (Autocorrelation in FE is present)
No Serial Autocorrelation (FD Model)	$t = -7.1$	$P = 0\%$	Reject Null (Autocorrelation in FD is present)
Constant Returns to Scale ($\beta_L + \beta_K = 1$)	$W = 37.35$	$P = 0\%$	Reject Null (No Constant Returns to Scale)

Tabel 3: The table displays test statistics and p-values for strict exogeneity, serial autocorrelation, and constant returns to scale.

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 $\alpha = 0$

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

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:

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

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.

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

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:

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

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:

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

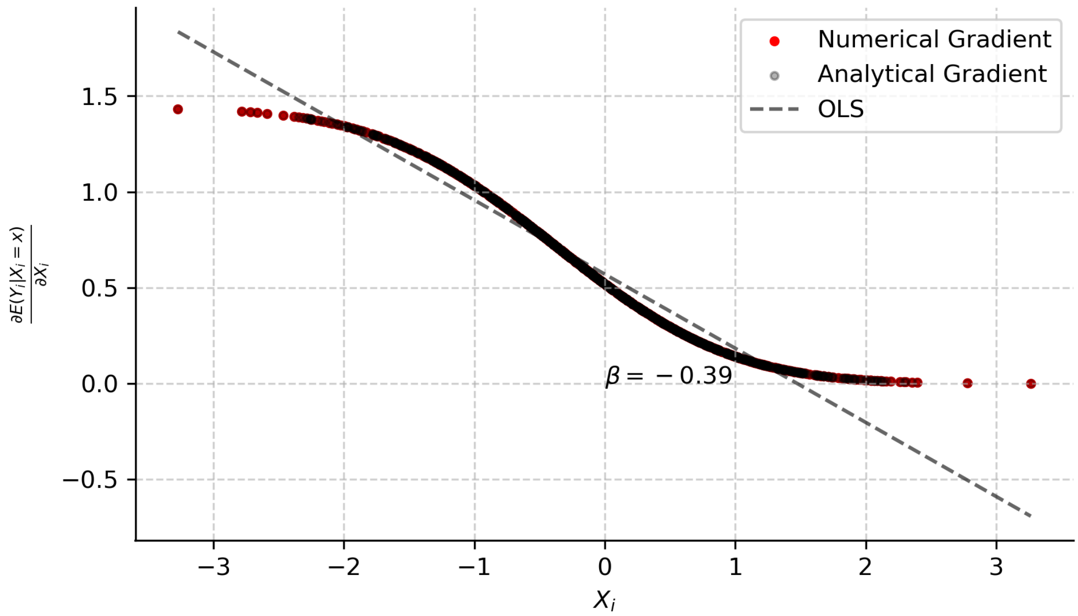
as required, with 9.5 following directly from the symmetry property.

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 $\widetilde{\text{Avar}}(\hat{\theta})$.