Interval Regression with Sample Selection

Arne Henningsen, Sebastian Petersen, Géraldine Henningsen

May 3, 2019

This 'vignette' is largely based on Petersen et al. (2017).

1 Model Specification

The general specification of an interval regression model with sample selection is:

$$y_i^{S*} = \boldsymbol{\beta}^{S\prime} \boldsymbol{x}_i^S + \varepsilon_i^S \tag{1}$$

$$y_i^S = \begin{cases} 0 & \text{if } y_i^{S*} \leq 0 \\ 1 & \text{otherwise} \end{cases} \tag{2}$$

$$y_i^{O*} = \boldsymbol{\beta}^{O\prime} \boldsymbol{x}_i^O + \varepsilon_i^O \tag{3}$$

$$y_i^O = \begin{cases} \text{unknown} & \text{if } y_i^S = 0 \\ 1 & \text{if } \alpha_1 < y_i^{O*} \leq \alpha_2 \text{ and } y_i^S = 1 \\ 2 & \text{if } \alpha_2 < y_i^{O*} \leq \alpha_3 \text{ and } y_i^S = 1 \\ \vdots & \\ M & \text{if } \alpha_M < y_i^{O*} \leq \alpha_{M+1} \text{ and } y_i^S = 1 \end{cases} \tag{4}$$

$$\left(\begin{array}{c} \varepsilon_i^S \\ \varepsilon_i^O \end{array} \right) \sim N_2 \left(\left(\begin{array}{c} 0 \\ 0 \end{array} \right), \begin{bmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{bmatrix} \right), \tag{5}$$

where subscript i indicates the observation, y_i^{O*} is a latent outcome variable, y_i^O is a partially observed categorical variable that indicates in which interval y_i^{O*} lies, M is the number of intervals, $\alpha_1, \ldots, \alpha_{M+1}$ are the boundaries of the intervals (whereas frequently but not necessarily $\alpha_1 = -\infty$ and $\alpha_{M+1} = \infty$), y_i^S is a binary variable that indicates whether y_i^O is observed, y_i^{S*} is a latent variable that indicates the "tendency" that y_i^S is one, $\boldsymbol{x}_i^S$ and $\boldsymbol{x}_i^O$ are (column) vectors of explanatory variables for the selection equation and outcome equation, respectively, ε_i^S and ε_i^O are random disturbance terms that have a joint bivariate normal distribution, and $\boldsymbol{\beta}^S$ and $\boldsymbol{\beta}^O$ are (column) vectors and ρ and σ are scalars of unknown model parameters.

2 Log-Likelihood Function

The probability that y_i^O is unobserved is:

$$P\left(y_i^S = 0\right) = P\left(y_i^{S*} \leq 0\right) \tag{6}$$
$$= P\left(\boldsymbol{\beta}^{S\prime} \boldsymbol{x}_i^S + \varepsilon_i^S \leq 0\right) \tag{7}$$
$$= P\left(\varepsilon_i^S \leq -\boldsymbol{\beta}^{S\prime} \boldsymbol{x}_i^S\right) \tag{8}$$

The probability that y_i^O is observed and indicates that y_i^{O*} lies in the mth interval is:

$$P\left(y_i^S = 1 \wedge y_i^O = m\right) = P\left(y_i^{S*} > 0 \wedge \alpha_m < y_i^{O*} \leq \alpha_{m+1}\right) \tag{9}$$
$$= P\left(\boldsymbol{\beta}^{S\prime} \boldsymbol{x}_i^S + \varepsilon_i^S > 0 \wedge \alpha_m < \boldsymbol{\beta}^{O\prime} \boldsymbol{x}_i^O + \varepsilon_i^O \leq \alpha_{m+1}\right) \tag{10}$$
$$= P\left(\varepsilon_i^S > -\boldsymbol{\beta}^{S\prime} \boldsymbol{x}_i^S \wedge \alpha_m - \boldsymbol{\beta}^{O\prime} \boldsymbol{x}_i^O < \varepsilon_i^O \leq \alpha_{m+1} - \boldsymbol{\beta}^{O\prime} \boldsymbol{x}_i^O\right) \tag{11}$$

The log-likelihood contribution of the ith observation is:

$$\begin{aligned}
\ell_i =&(1 - y_i^S) \ln\left[\Phi\left(-\boldsymbol{\beta}^{S\prime} \boldsymbol{x}_i^S\right)\right] \qquad (12)\\
&+ \sum_{m=1}^{M} y_i^S (y_i^O = m) \ln\left[\Phi_2\left(\frac{\alpha_{m+1} - \boldsymbol{\beta}^{O\prime} \boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime} \boldsymbol{x}_i^S, -\rho\right)\right.\\
&\left. -\Phi_2\left(\frac{\alpha_m - \boldsymbol{\beta}^{O\prime} \boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime} \boldsymbol{x}_i^S, -\rho\right)\right],
\end{aligned}$$

where $\Phi(.)$ indicates the cumulative distribution function of the univariate standard normal distribution and $\Phi_2(.)$ indicates the cumulative distribution function of the bivariate standard normal distribution.

3 Restricting coefficients ρ and σ^2

The parameter ρ needs to be in the interval $(-1, 1)$. In order to restrict ρ to be in this interval, we estimate $\arctan(\rho)$ instead of ρ so that the derived parameter $\rho = \tan(\arctan(\rho))$ is always in the interval $(-1, 1)$. We use the delta method to calculate approximate standard errors of the derived parameter ρ, whereas the corresponding element of the Jacobian matrix is:

$$\frac{\partial \tan(\arctan(\rho))}{\partial \arctan(\rho)} = \frac{\partial \rho}{\partial \arctan(\rho)} = (1 + \rho^2) \tag{13}$$

The parameter σ needs to be strictly positive, i.e. $\sigma > 0$. In order to restrict σ to be strictly positive, we estimate $\log(\sigma)$ instead of σ or σ^2 so that the derived parameters $\sigma = \exp(\log(\sigma))$ and $\sigma^2 = \exp(2\ \log(\sigma))$ are always strictly positive. We use the delta method to calculate approximate standard errors of the derived parameters σ and σ^2, whereas the corresponding elements of the Jacobian matrix are:

$$\frac{\partial \exp(\log(\sigma))}{\partial \log(\sigma)} = \exp(\log(\sigma)) = \sigma \tag{14}$$

$$\frac{\partial \exp(2\ \log(\sigma))}{\partial \log(\sigma)} = 2\ \exp(2\ \log(\sigma)) = 2\ \sigma^2 \tag{15}$$

4 Gradients of the CDF of the bivariate standard normal distribution

In order to facilitate the calculation of the gradients of the log-likelihood function, we calculate the partial derivatives of the cumulative distribution function (CDF) of the bivariate standard normal distribution:

$$\Phi_2(x_1, x_2, \rho) = \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} \phi_2(a_1, a_2, \rho)\ da_1\ da_2, \tag{16}$$

where $\phi_2(.)$ is the probability density function (PDF) of the bivariate standard normal distribution:

$$\phi_2(x_1, x_2, \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \cdot \exp\left(-\frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{2(1-\rho^2)}\right) \tag{17}$$

In the following, we check equation (17) by a simple numerical example:

```
> library( "mvtnorm" )
> library( "maxLik" )
> x1 <- 0.4
> x2 <- -0.3
> rho <- -0.6
> sigma <-  matrix( c( 1, rho, rho, 1 ), nrow = 2 )
> dens <- dmvnorm( c( x1, x2 ), sigma = sigma )
> print( dens )

[1] 0.1831324

> all.equal( dens, ( 2 * pi * sqrt( 1 - rho^2 ) )^(-1) *
+    exp( - ( x1^2 - 2 * rho * x1 * x2 + x2^2 ) / ( 2 * ( 1 - rho^2 ) ) ) )

[1] TRUE
```

4.1 Gradients with respect to the limits (x_1 and x_2)

$$\frac{\partial \Phi_2(x_1, x_2, \rho)}{\partial x_2} = \int_{-\infty}^{x_1} \phi_2(a_1, x_2, \rho) \; da_1 \tag{18}$$
$$= \int_{-\infty}^{x_1} \phi(a_1 | x_2, \rho) \phi(x_2) \; da_1 \tag{19}$$
$$= \int_{-\infty}^{x_1} \tilde{\phi}\left(a_1, \rho x_2, 1 - \rho^2\right) \phi(x_2) \; da_1 \tag{20}$$
$$= \int_{-\infty}^{x_1} \phi\left(\frac{a_1 - \rho x_2}{\sqrt{1 - \rho^2}}\right) \left(\sqrt{1 - \rho^2}\right)^{-1} \phi(x_2) \; da_1 \tag{21}$$
$$= \int_{-\infty}^{x_1} \phi\left(\frac{a_1 - \rho x_2}{\sqrt{1 - \rho^2}}\right) \left(\sqrt{1 - \rho^2}\right)^{-1} da_1 \; \phi(x_2) \tag{22}$$
$$= \int_{-\infty}^{\frac{x_1 - \rho x_2}{\sqrt{1 - \rho^2}}} \phi(a_1) \; da_1 \; \phi(x_2) \tag{23}$$
$$= \Phi\left(\frac{x_1 - \rho x_2}{\sqrt{1 - \rho^2}}\right) \phi(x_2), \tag{24}$$

where $\tilde{\phi}(\,, \mu, \sigma^2)$ indicates the density function of a normal distribution with mean μ and variance σ^2.

In the following, we use the same simple numerical example as in the beginning of section 4 to check the above derivations. First, we check whether the PDF of the bivariate standard normal distribution, i.e. $\phi_2(x_1, x_2, \rho)$ (part of equation 18), is equal to $\tilde{\phi}\left(x_1, \rho x_2, 1 - \rho^2\right) \phi(x_2)$ (part of equation 20) and equal to $\phi\left((x_1 - \rho x_2)/(\sqrt{1 - \rho^2})\right)$ $\left(\sqrt{1 - \rho^2}\right)^{-1} \phi(x_2)$ (part of equations 21 and 22):

```
> all.equal( dens, dnorm( x1, rho * x2, sqrt( 1 - rho^2 ) ) * dnorm(x2) )

[1] TRUE

> all.equal( dens, ( dnorm( ( x1 - rho * x2 ) / sqrt( 1 - rho^2 ) ) /
+    sqrt( 1 - rho^2 ) ) * dnorm(x2) )

[1] TRUE
```

In the following, we will numerically calculate the derivative of the cumulative distribution function of the bivaraite normal distribution (equation 16) with respect to x_2 and check wehther this partial derivative is equal to the right-hand sides of equations (18), (21), (22), and (24):

```
> funX2 <- function( a2 ) {
+    prob <- pmvnorm( upper = c( x1, a2 ), sigma = sigma )
```

```
+    return( prob )
+ }
> grad <- c( numericGradient( funX2, x2 ) )
> print( grad )
```

```
[1] 0.2320142
```

```
> funX1 <- function( a1 ) {
+    dens <- rep( NA, length( a1 ) )
+    for( i in 1:length( a1 ) ) {
+       dens[i] <- dmvnorm( c( a1[i], x2 ), sigma = sigma )
+    }
+    return( dens )
+ }
> all.equal( grad, integrate( funX1, lower = -Inf, upper = x1 )$value )
```

```
[1] TRUE
```

```
> funX1a <- function( a1 ) {
+    dens <- rep( NA, length( a1 ) )
+    for( i in 1:length( a1 ) ) {
+       dens[i] <- ( dnorm( ( a1[i] - rho * x2 ) / sqrt( 1 - rho^2 ) ) /
+             sqrt(1-rho^2) ) * dnorm(x2)
+    }
+    return( dens )
+ }
> all.equal( grad, integrate( funX1a, lower = -Inf, upper = x1 )$value )
```

```
[1] TRUE
```

```
> funX1b <- function( a1 ) {
+    dens <- rep( NA, length( a1 ) )
+    for( i in 1:length( a1 ) ) {
+       dens[i] <- dnorm( ( a1[i] - rho * x2 ) / sqrt( 1 - rho^2 ) ) /
+          sqrt(1-rho^2)
+    }
+    return( dens )
+ }
> all.equal( grad,
+    integrate( funX1b, lower = -Inf, upper = x1 )$value * dnorm(x2) )
```

```
[1] TRUE
```

```
> all.equal( grad,
+    pnorm( ( x1 - rho * x2 ) / sqrt( 1 - rho^2 ) ) * dnorm( x2 ) )
```

```
[1] TRUE
```

4.2 Gradients with respect to the coefficient of correlation (ρ)

$$\frac{\partial \Phi_2(x_1, x_2, \rho)}{\partial \rho} \tag{25}$$

$$= \frac{\partial \left[\int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \phi_2(a_1, a_2, \rho)\, da_2\, da_1\right]}{\partial \rho} \tag{26}$$

$$= \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \frac{\partial \phi_2(a_1, a_2, \rho)}{\partial \rho}\, da_2\, da_1 \tag{27}$$

$$= \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \frac{\partial}{\partial \rho} \left(\frac{\exp\left(-\frac{a_1^2 - 2\rho a_1 a_2 + a_2^2}{2(1-\rho^2)}\right)}{2\pi\sqrt{1-\rho^2}} \right) da_2\, da_1 \tag{28}$$

$$= \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \frac{1}{2\pi} \frac{\partial}{\partial \rho} \left(\frac{\exp\left(-\frac{a_1^2 - 2\rho a_1 a_2 + a_2^2}{2(1-\rho^2)}\right)}{\sqrt{1-\rho^2}} \right) da_2\, da_1 \tag{29}$$

$$= \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \frac{1}{2\pi} \left(\frac{\frac{\partial}{\partial \rho}\left(\exp\left(-\frac{a_1^2 - 2\rho a_1 a_2 + a_2^2}{2(1-\rho^2)}\right)\right) \cdot \sqrt{1-\rho^2}}{1-\rho^2} - \frac{\frac{\partial}{\partial \rho}(\sqrt{1-\rho^2}) \cdot \exp\left(-\frac{a_1^2 - 2\rho a_1 a_2 + a_2^2}{2(1-\rho^2)}\right)}{1-\rho^2} \right) da_2\, da_1 \tag{30}$$

$$= \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \frac{1}{2\pi} \left(\frac{\frac{\partial}{\partial \rho}\left(-\frac{a_1^2 - 2\rho a_1 a_2 + a_2^2}{2(1-\rho^2)}\right) \cdot \exp\left(-\frac{a_1^2 - 2\rho a_1 a_2 + a_2^2}{2(1-\rho^2)}\right) \cdot \sqrt{1-\rho^2}}{1-\rho^2} - \frac{\left(-\frac{\rho}{\sqrt{1-\rho^2}}\right) \cdot \exp\left(-\frac{a_1^2 - 2\rho a_1 a_2 + a_2^2}{2(1-\rho^2)}\right)}{1-\rho^2} \right) da_2\, da_1 \tag{31}$$

$$=\int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\frac{1}{2\pi}\left(\frac{\frac{\left(-4\rho(a_1^2-2\rho a_1a_2+a_2^2)-2(1-\rho^2)(-2a_1a_2)\right)}{4(1-\rho^2)^2}\cdot\exp\left(-\frac{a_1^2-2\rho a_1a_2+a_2^2}{2(1-\rho^2)}\right)\cdot\sqrt{1-\rho^2}}{1-\rho^2}\right. \tag{32}$$

$$\left.-\frac{\left(-\frac{\rho}{\sqrt{1-\rho^2}}\right)\cdot\exp\left(-\frac{a_1^2-2\rho a_1a_2+a_2^2}{2(1-\rho^2)}\right)}{1-\rho^2}\right)\,da_2\,da_1$$

$$=\int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\frac{1}{2\pi}\left(\frac{\frac{\left(-4\rho(a_1^2-2\rho a_1a_2+a_2^2)-2(1-\rho^2)(-2a_1a_2)\right)}{4(1-\rho^2)^2}\cdot\sqrt{1-\rho^2}}{1-\rho^2}-\frac{\left(-\frac{\rho}{\sqrt{1-\rho^2}}\right)}{1-\rho^2}\right) \tag{33}$$

$$\cdot\exp\left(-\frac{a_1^2-2\rho a_1a_2+a_2^2}{2(1-\rho^2)}\right)\,da_2\,da_1$$

$$=\int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\frac{1}{2\pi}\left(\frac{\left(-4\rho(a_1^2-2\rho a_1a_2+a_2^2)-2(1-\rho^2)(-2a_1a_2)\right)}{4(1-\rho^2)^{\frac{5}{3}}}+\frac{\rho}{(1-\rho^2)^{\frac{3}{2}}}\right) \tag{34}$$
$$\cdot\exp\left(-\frac{a_1^2-2\rho a_1a_2+a_2^2}{2(1-\rho^2)}\right)\,da_2\,da_1$$

$$=\int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\frac{1}{2\pi}\left(\frac{\left(-4\rho(a_1^2-2\rho a_1a_2+a_2^2)-2(1-\rho^2)(-2a_1a_2)\right)}{4(1-\rho^2)^{\frac{5}{2}}}+\frac{\rho}{(1-\rho^2)^{\frac{3}{2}}}\right) \tag{35}$$
$$\cdot\exp\left(-\frac{a_1^2-2\rho a_1a_2+a_2^2}{2(1-\rho^2)}\right)\,da_2\,da_1$$

$$=\int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\frac{1}{2\pi}\left(\frac{\rho}{(1-\rho^2)^{\frac{3}{2}}}-\frac{\rho(a_1^2-\rho a_1a_2+a_2^2)-a_1a_2}{(1-\rho^2)^{\frac{5}{2}}}\right) \tag{36}$$
$$\cdot\exp\left(-\frac{a_1^2-2\rho a_1a_2+a_2^2}{2(1-\rho^2)}\right)\,da_2\,da_1$$

$$=\frac{1}{2\pi\sqrt{(1-\rho^2)}}\int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\left(\frac{\rho}{1-\rho^2}-\frac{\rho(a_1^2-\rho a_1a_2+a_2^2)-a_1a_2}{(1-\rho^2)^2}\right) \tag{37}$$
$$\cdot\exp\left(-\frac{a_1^2-2\rho a_1a_2+a_2^2}{2(1-\rho^2)}\right)\,da_2\,da_1$$

$$=\frac{1}{2\pi\sqrt{(1-\rho^2)}}\int_{-\infty}^{x_1}\left|\left(-\frac{2a_1-2\rho a_2}{2(1-\rho^2)}\right)\cdot\exp\left(-\frac{a_1^2-2\rho a_1 a_2+a_2^2}{2(1-\rho^2)}\right)\right|_{-\infty}^{x_2} da_1 \tag{38}$$

$$=\frac{1}{2\pi\sqrt{(1-\rho^2)}}\int_{-\infty}^{x_1}\left(\left(-\frac{2a_1-2\rho x_2}{2(1-\rho^2)}\right)\cdot\exp\left(-\frac{a_1^2-2\rho a_1 x_2+x_2^2}{2(1-\rho^2)}\right)\right. \tag{39}$$

$$\left.-\lim_{a_2\to-\infty}\frac{1}{2(1-\rho^2)}\frac{-2a_1+2\rho a_2}{\exp\left(\frac{a_1^2-2\rho a_1 a_2+a_2^2}{2(1-\rho^2)}\right)}\right) da_1$$

Applying L'Hospital on the last term leads to

$$=\frac{1}{2\pi\sqrt{(1-\rho^2)}}\int_{-\infty}^{x_1}\left(\left(-\frac{2a_1-2\rho x_2}{2(1-\rho^2)}\right)\cdot\exp\left(-\frac{a_1^2-2\rho a_1 x_2+x_2^2}{2(1-\rho^2)}\right)-0\right) da_1 \tag{40}$$

$$=\frac{1}{2\pi\sqrt{(1-\rho^2)}}\int_{-\infty}^{x_1}\left(-\frac{2a_1-2\rho x_2}{2(1-\rho^2)}\right)\cdot\exp\left(-\frac{a_1^2-2\rho a_1 x_2+x_2^2}{2(1-\rho^2)}\right) da_1 \tag{41}$$

$$=\frac{1}{2\pi\sqrt{(1-\rho^2)}}\left|\exp\left(-\frac{a_1^2-2\rho a_1 x_2+x_2^2}{2(1-\rho^2)}\right)\right|_{-\infty}^{x_1} \tag{42}$$

$$=\frac{1}{2\pi\sqrt{(1-\rho^2)}}\cdot\exp\left(-\frac{x_1^2-2\rho x_1 x_2+x_2^2}{2(1-\rho^2)}\right) \tag{43}$$

$$=\phi_2(x_1,x_2,\rho) \tag{44}$$

This result is in line with Sibuya (1960) and Sungur (1990).

In the following, we will numerically calculate the derivative of the cumulative distribution function of the bivariate normal distribution (equation 26) with respect to ρ and check whether this partial derivative is equal to the right-hand sides of equation (44):

```
> # Numerical gradient of the PDF w.r.t. rho
> funrho <- function( p ) {
+    prob <- dmvnorm( x = c( x1, x2 ),
+       sigma = matrix( c( 1, p, p, 1 ), nrow = 2 ) )
+    return( prob )
+ }
> grad <- c( numericGradient( funrho, rho ) )
> print( grad )
```

```
[1] -0.1775883
```

```
> # Comparison with analytical gradient for rho
> efun <- exp(-(x1^2 - 2 * rho * x1 * x2 + x2^2)/(2*(1 - rho^2)))
```

```
> all.equal( grad,
+ ( (-((2*rho*(-2*rho*x1*x2+x1^2+x2^2) - 2*x1*x2*(1-rho^2)) * efun)/
+     (2*(1-rho^2)^(3/2) )) +
+     ((rho*efun)/(sqrt(1-rho^2))) ) /
+     (2*pi*(1-rho^2)) )
```

```
[1] TRUE
```

```
[1] TRUE
```

```
[1] TRUE
```

```
[1] TRUE
```

```
> # Numerical gradient of the CDF w.r.t. rho
> cdfRho <- function( p, xa = x1, xb = x2 ) {
+    prob <- pmvnorm( upper = c( xa, xb ),
+       sigma = matrix( c( 1, p, p, 1 ), nrow = 2 ) )
+    return( prob )
+ }
> grad <- c( numericGradient( cdfRho, rho ) )
> print( grad )
```

```
[1] 0.1831324
```

```
> # comparison with analytical gradient
> all.equal( grad, dmvnorm( x = c( x1, x2 ),
+       sigma = matrix( c( 1, rho, rho, 1 ), nrow = 2 ) ) )
```

```
[1] TRUE
```

```
> # comparisons with other values
> compDerivRho <- function( xa, xb, p ) {
+    dn <- c( numericGradient( cdfRho, p, xa = xa, xb = xb ) )
+    da <- dmvnorm( x = c( xa, xb ),
+       sigma = matrix( c( 1, p, p, 1 ), nrow = 2 ) )
+    return( all.equal( dn, da ) )
+ }
> compDerivRho( x1, x2, rho )
```

```
[1] TRUE
```

```
> compDerivRho( 0.5, x2, rho )
```

```
[1] TRUE
```

```
> compDerivRho( 2.5, x2, rho )
[1] TRUE
> compDerivRho( x1, -2, rho )
[1] TRUE
> compDerivRho( x1, x2, 0.2 )
[1] TRUE
> compDerivRho( x1, x2, 0.98 )
[1] TRUE
```

5 Gradients of the Log-Likelihood Function

5.1 Gradients with respect to the parameters of the selection equation (β^S)

First, we use equation (24), to determine the derivative of the bivariate standard normal distribution with respect to the parameter $\boldsymbol{\beta}^S$ as part of the loglikelihood function:

$$\frac{\partial \Phi_2\left(\frac{\alpha_m - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S, -\rho\right)}{\partial \beta^S} = \Phi\left(\frac{\frac{\alpha_m - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma} - \rho\boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S}{\sqrt{1-\rho^2}}\right)\phi(\boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S) \cdot \frac{\partial \boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S}{\partial \boldsymbol{\beta}^S} \tag{45}$$

$$= \Phi\left(\frac{\frac{\alpha_m - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma} + \rho\boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S}{\sqrt{1-\rho^2}}\right)\phi(\boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S) \cdot \boldsymbol{x}_i^S \tag{46}$$

Using this result we can now derive the gradient for $\boldsymbol{\beta}^S$ in the log-likelihood function:

$$\frac{\partial \ell_i}{\partial \beta^S} = \frac{\partial}{\partial \beta^S}\Bigg((1-y_i^S)\ln\left[\Phi\left(-\boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S\right)\right] \tag{47}$$

$$+ \sum_{m=1}^{M} y_i^S(y_i^O = m)\ln\Bigg[\Phi_2\left(\frac{\alpha_{m+1} - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S, -\rho\right)$$

$$-\Phi_2\left(\frac{\alpha_m - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S, -\rho\right)\Bigg]\Bigg)$$

$$= (1-y_i^S)\frac{\partial}{\partial \beta^S}\left(\ln\left[\Phi\left(-\boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S\right)\right]\right) \tag{48}$$

$$+ \sum_{m=1}^{M} y_i^S(y_i^O = m)\frac{\partial}{\partial \beta^S}\Bigg(\ln\Bigg[\Phi_2\left(\frac{\alpha_{m+1} - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S, -\rho\right)$$

$$-\Phi_2\left(\frac{\alpha_m - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S, -\rho\right)\Bigg]\Bigg)$$

$$= (1-y_i^S)\frac{\phi\left(-\boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S\right)\cdot(-\boldsymbol{x}_i^S)}{\Phi\left(-\boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S\right)} \tag{49}$$

$$+ \sum_{m=1}^{M} y_i^S(y_i^O = m)\frac{\frac{\partial \Phi_2\left(\frac{\alpha_{m+1} - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S, -\rho\right)}{\partial \beta^S} - \frac{\partial \Phi_2\left(\frac{\alpha_m - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S, -\rho\right)}{\partial \beta^S}}{\Phi_2\left(\frac{\alpha_{m+1} - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S, -\rho\right) - \Phi_2\left(\frac{\alpha_m - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S, -\rho\right)}$$

$$= (1-y_i^S)\frac{\phi\left(-\boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S\right)\cdot(-\boldsymbol{x}_i^S)}{\Phi\left(-\boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S\right)} \tag{50}$$

$$
\begin{aligned}
&+\sum_{m=1}^{M} y_i^S (y_i^O = m) \frac{1}{\Phi_2\left(\frac{\alpha_{m+1}-\boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S, -\rho\right) - \Phi_2\left(\frac{\alpha_{m}-\boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S, -\rho\right)} \\
&\quad \left(\Phi\left(\frac{\frac{\alpha_{m+1}-\boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma} + \rho\boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S}{\sqrt{1-\rho^2}}\right)\phi\left(\boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S\right)\cdot \boldsymbol{x}_i^S \right. \\
&\qquad \left. -\ \Phi\left(\frac{\frac{\alpha_{m}-\boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma} + \rho\boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S)}{\sqrt{1-\rho^2}}\right)\phi\left(\boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S\right)\cdot \boldsymbol{x}_i^S \right) \\
&= (1-y_i^S)\frac{\phi\left(-\boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S\right)\cdot(-\boldsymbol{x}_i^S)}{\Phi\left(-\boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S\right)} \qquad (51) \\
&+\sum_{m=1}^{M} y_i^S (y_i^O = m) \frac{\left(\Phi\left(\frac{\frac{\alpha_{m+1}-\boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}+\rho\boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S}{\sqrt{1-\rho^2}}\right) - \Phi\left(\frac{\frac{\alpha_{m}-\boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}+\rho\boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S}{\sqrt{1-\rho^2}}\right)\right)\phi\left(\boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S\right)\cdot\boldsymbol{x}_i^S}{\Phi_2\left(\frac{\alpha_{m+1}-\boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S, -\rho\right) - \Phi_2\left(\frac{\alpha_{m}-\boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S, -\rho\right)}
\end{aligned}
$$

5.2 Gradients with respect to the parameters in the outcome equation (β^O)

Analogous to $\boldsymbol{\beta}^S$ and by using equation (24) we derive the gradient of $\boldsymbol{\beta}^O$:

$$\frac{\partial \Phi_2\left(\frac{\alpha_m-\boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S, -\rho\right)}{\partial \beta^O} = \Phi\left(\frac{\boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S + \rho\frac{\alpha_m-\boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}}{\sqrt{1-\rho^2}}\right)\phi\left(\frac{\alpha_m-\boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}\right)\cdot\left(-\frac{\boldsymbol{x}_i^O}{\sigma}\right) \tag{52}$$

Using this result we derive the gradient for the outcome parameter $\boldsymbol{\beta}^O$ for the log-likelihood function:

$$\begin{aligned}
\frac{\partial \ell_i}{\partial \beta^O} = & \frac{\partial}{\partial \beta^O}\Bigg((1-y_i^S)\ln\left[\Phi\left(-\boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S\right)\right] && (53)\\
& + \sum_{m=1}^{M} y_i^S(y_i^O=m)\ln\Bigg[\Phi_2\left(\frac{\alpha_{m+1}-\boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S, -\rho\right) \\
& -\Phi_2\left(\frac{\alpha_m-\boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S, -\rho\right)\Bigg]\Bigg) \\
= & \sum_{m=1}^{M} y_i^S(y_i^O=m)\frac{\partial}{\partial \beta^O}\Bigg(\ln\Bigg[\Phi_2\left(\frac{\alpha_{m+1}-\boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S, -\rho\right) && (54)\\
& -\Phi_2\left(\frac{\alpha_m-\boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S, -\rho\right)\Bigg]\Bigg) \\
= & \sum_{m=1}^{M} y_i^S(y_i^O=m)\frac{\frac{\partial \Phi_2\left(\frac{\alpha_{m+1}-\boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S, -\rho\right)}{\partial \beta^O} - \frac{\partial \Phi_2\left(\frac{\alpha_m-\boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S, -\rho\right)}{\partial \beta^O}}{\Phi_2\left(\frac{\alpha_{m+1}-\boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S, -\rho\right) - \Phi_2\left(\frac{\alpha_m-\boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S, -\rho\right)} && (55)\\
= & \sum_{m=1}^{M} y_i^S(y_i^O=m)\cdot\frac{1}{\Phi_2\left(\frac{\alpha_{m+1}-\boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S, -\rho\right) - \Phi_2\left(\frac{\alpha_m-\boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S, -\rho\right)} && (56)\\
& \Bigg(\Phi\left(\frac{\boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S + \rho\frac{\alpha_{m+1}-\boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}}{\sqrt{1-\rho^2}}\right)\phi\left(\frac{\alpha_{m+1}-\boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}\right)\cdot\left(-\frac{\boldsymbol{x}_i^O}{\sigma}\right) \\
& -\Phi\left(\frac{\boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S + \rho\frac{\alpha_m-\boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}}{\sqrt{1-\rho^2}}\right)\phi\left(\frac{\alpha_m-\boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}\right)\cdot\left(-\frac{\boldsymbol{x}_i^O}{\sigma}\right)\Bigg)
\end{aligned}$$

$$= \sum_{m=1}^{M} y_i^S (y_i^O = m) \cdot \frac{1}{\Phi_2\left(\frac{\alpha_{m+1} - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S, -\rho\right) - \Phi_2\left(\frac{\alpha_m - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S, -\rho\right)} \tag{57}$$

$$\left(\Phi\left(\frac{\boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S + \rho\frac{\alpha_{m+1} - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}}{\sqrt{1-\rho^2}}\right)\phi\left(\frac{\alpha_{m+1} - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}\right)\right.$$
$$\left. -\Phi\left(\frac{\boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S + \rho\frac{\alpha_m - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}}{\sqrt{1-\rho^2}}\right)\phi\left(\frac{\alpha_m - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}\right)\right) \cdot \left(-\frac{\boldsymbol{x}_i^O}{\sigma}\right)$$

5.3 Gradients with respect to the coefficient of correlation (ρ)

Given the result that the derivative of the CDF with respect to ρ is equal to the PDF (see equation 44), we can also derive the gradient of the correlation parameter (ρ):

$$\frac{\partial \ell_i}{\partial \rho} = \frac{\partial}{\partial \rho}\left((1 - y_i^S)\ln\left[\Phi\left(-\boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S\right)\right]\right. \tag{58}$$
$$+ \sum_{m=1}^{M} y_i^S (y_i^O = m) \ln\left[\Phi_2\left(\frac{\alpha_{m+1} - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S, -\rho\right)\right.$$
$$\left.\left. -\Phi_2\left(\frac{\alpha_m - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S, -\rho\right)\right]\right)$$
$$= \sum_{m=1}^{M} y_i^S (y_i^O = m) \frac{\partial}{\partial \rho}\left(\ln\left[\Phi_2\left(\frac{\alpha_{m+1} - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S, -\rho\right)\right.\right. \tag{59}$$
$$\left.\left. -\Phi_2\left(\frac{\alpha_m - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S, -\rho\right)\right]\right)$$
$$= \sum_{m=1}^{M} y_i^S (y_i^O = m) \frac{\phi_2\left(\frac{\alpha_m - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S, -\rho\right) - \phi_2\left(\frac{\alpha_{m+1} - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S, -\rho\right)}{\Phi_2\left(\frac{\alpha_{m+1} - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S, -\rho\right) - \Phi_2\left(\frac{\alpha_m - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S, -\rho\right)} \tag{60}$$

$$\frac{\partial \ell_i}{\partial \operatorname{arctanh}(\rho)} = \frac{\partial \ell_i}{\partial \rho}\frac{\partial \rho}{\partial \operatorname{arctanh}(\rho)} = \frac{\partial \ell_i}{\partial \rho}\,(1 - \rho^2) \tag{61}$$

5.4 Gradients with respect to the standard deviation used for normalisation (σ)

Finally, we derive the gradient for σ in the same way as we did for β^S and β^O:

$$
\begin{aligned}
&\frac{\partial \Phi_2\left(\frac{\alpha_m - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S, -\rho\right)}{\partial \sigma} \\
&= \Phi\left(\frac{\boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S + \rho\frac{\alpha_m - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}}{\sqrt{1-\rho^2}}\right)\phi\left(\frac{\alpha_m - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}\right)\cdot\frac{\boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O - \alpha_m}{\sigma^2} && (62)
\end{aligned}
$$

$$
\begin{aligned}
&\lim_{\alpha_m \to \infty} \frac{\partial \Phi_2\left(\frac{\alpha_m - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S, -\rho\right)}{\partial \sigma} \\
&= \lim_{\alpha_m \to \infty} \Phi\left(\frac{\boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S + \rho\frac{\alpha_m - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}}{\sqrt{1-\rho^2}}\right)\phi\left(\frac{\alpha_m - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}\right)\cdot\frac{\boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O - \alpha_m}{\sigma^2} && (63) \\
&= \Phi\left(\frac{\boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S + \rho\frac{\alpha_m - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}}{\sqrt{1-\rho^2}}\right)\lim_{\alpha_m \to \infty}\phi\left(\frac{\alpha_m - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}\right)\cdot\frac{\boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O - \alpha_m}{\sigma^2} && (64) \\
&= \Phi\left(\frac{\boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S + \rho\frac{\alpha_m - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}}{\sqrt{1-\rho^2}}\right)\lim_{\alpha_m \to \infty}\frac{\frac{\boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O - \alpha_m}{\sigma^2}}{\left(\phi\left(\frac{\alpha_m - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}\right)\right)^{-1}} && (65) \\
&= \Phi\left(\frac{\boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S + \rho\frac{\alpha_m - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}}{\sqrt{1-\rho^2}}\right) && (66) \\
&\qquad \lim_{\alpha_m \to \infty}\frac{-\frac{1}{\sigma^2}}{-\left(\phi\left(\frac{\alpha_m - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}\right)\right)^{-2}\left(-\frac{\alpha_m - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}\right)\phi\left(\frac{\alpha_m - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}\right)\frac{1}{\sigma}} \\
&= \Phi\left(\frac{\boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S + \rho\frac{\alpha_m - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}}{\sqrt{1-\rho^2}}\right)\lim_{\alpha_m \to \infty}\frac{-\frac{1}{\sigma}}{\left(\phi\left(\frac{\alpha_m - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}\right)\right)^{-1}\frac{\alpha_m - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}} && (67) \\
&= \Phi\left(\frac{\boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S + \rho\frac{\alpha_m - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}}{\sqrt{1-\rho^2}}\right)\lim_{\alpha_m \to \infty}\frac{-\frac{1}{\sigma}\phi\left(\frac{\alpha_m - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}\right)}{\frac{\alpha_m - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}} && (68) \\
&= 0 && (69)
\end{aligned}
$$

Similarly:

$$
\lim_{\alpha_m \to -\infty} \frac{\partial \Phi_2\left(\frac{\alpha_m - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S, -\rho\right)}{\partial \sigma} = 0 \qquad (70)
$$

$$
\begin{aligned}
\frac{\partial \ell_i}{\partial \sigma} = \frac{\partial}{\partial \sigma}\Bigg(& (1 - y_i^S) \ln\left[\Phi\left(-\boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S\right)\right] && (71)\\
& + \sum_{m=1}^{M} y_i^S (y_i^O = m) \ln\left[\Phi_2\left(\frac{\alpha_{m+1} - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S, -\rho\right)\right.\\
& \left.\left. -\Phi_2\left(\frac{\alpha_m - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S, -\rho\right)\right]\right)\\
= \sum_{m=1}^{M} y_i^S (y_i^O = m) \frac{\partial}{\partial \sigma}\Bigg(& \ln\left[\Phi_2\left(\frac{\alpha_{m+1} - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S, -\rho\right)\right. && (72)\\
& \left.\left. -\Phi_2\left(\frac{\alpha_m - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S, -\rho\right)\right]\right)
\end{aligned}
$$

$$
= \sum_{m=1}^{M} y_i^S (y_i^O = m) \frac{\frac{\partial \Phi_2\left(\frac{\alpha_{m+1} - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S, -\rho\right)}{\partial \sigma} - \frac{\partial \Phi_2\left(\frac{\alpha_m - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S, -\rho\right)\Big)}{\partial \sigma}}{\Phi_2\left(\frac{\alpha_{m+1} - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S, -\rho\right) - \Phi_2\left(\frac{\alpha_m - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S, -\rho\right)} \tag{73}
$$

$$
\begin{aligned}
= \sum_{m=1}^{M} & \frac{y_i^S (y_i^O = m)}{\Phi_2\left(\frac{\alpha_{m+1} - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S, -\rho\right) - \Phi_2\left(\frac{\alpha_m - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}, \boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S, -\rho\right)} && (74)\\
& \left(\Phi\left(\frac{\boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S + \rho\frac{\alpha_{m+1} - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}}{\sqrt{1-\rho^2}}\right) \phi\left(\frac{\alpha_{m+1} - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}\right) \cdot \frac{\boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O - \alpha_{m+1}}{\sigma^2}\right.\\
& \left. - \Phi\left(\frac{\boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S + \rho\frac{\alpha_m - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}}{\sqrt{1-\rho^2}}\right) \phi\left(\frac{\alpha_m - \boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sigma}\right) \cdot \frac{\boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O - \alpha_m}{\sigma^2}\right)
\end{aligned}
$$

$$
\frac{\partial \ell_i}{\partial \log(\sigma)} = \frac{\partial \ell_i}{\partial \sigma}\frac{\partial \sigma}{\partial \log(\sigma)} = \frac{\partial \ell_i}{\partial \sigma}\,\sigma \tag{75}
$$

6 Example with a Generated Dataset

7 Generate Dataset

```
> library( "mvtnorm" )
> # number of observations
> nObs <- 300
> # parameters
```

```
> betaS <- c( 1, 1, -1 )
> betaO <- c( 10, 4 )
> rho <- 0.4
> sigma <- 5
> # boundaries of the intervals
> bound <- c(-Inf,5,15,Inf)
> # set 'seed' of the pseudo random number generator
> # in order to always generate the same pseudo random numbers
> set.seed(123)
> # generate variables x1 and x2
> dat <- data.frame( x1 = rnorm( nObs ), x2 = rnorm( nObs ) )
> # variance-covariance matrix of the two error terms
> vcovMat <- matrix( c( 1, rho*sigma, rho*sigma, sigma^2 ), nrow = 2 )
> # generate the two error terms
> eps <- rmvnorm( nObs, sigma = vcovMat )
> dat$epsS <- eps[,1]
> dat$epsO <- eps[,2]
> # generate the selection variable
> dat$yS <- with( dat, betaS[1] + betaS[2] * x1 + betaS[3] * x2 + epsS ) > 0
> table( dat$yS )

FALSE  TRUE
   91   209

> # generate the unobserved/latent outcome variable
> dat$yOu <- with( dat, betaO[1] + betaO[2] * x1 + epsO )
> dat$yOu[ !dat$yS ] <- NA
> # obtain the intervals of the outcome variable
> dat$yO <- cut( dat$yOu, bound )
> table( dat$yO )

 (-Inf,5]    (5,15] (15, Inf]
       26       130        53
```

7.1 Estimation of the Model

In the following estimation, the starting values are obtained by a maximum-likelihood (ML) estimation of a tobit-2 model, where the dependent variable of the outcome equation is set to the mid points of the intervals:

```
> library( "sampleSelection" )
> res <- selection( yS ~ x1 + x2, yO ~ x1, data = dat, boundaries = bound )
> res

Call:
 selection(selection = yS ~ x1 + x2, outcome = yO ~ x1, data = dat,      boundaries = bou
```

```
Coefficients:
S:(Intercept)          S:x1          S:x2  O:(Intercept)          O:x1
       0.9820        0.9668       -1.2862        10.2403        2.6598
     logSigma     atanhRho         sigma        sigmaSq           rho
       1.6308        0.2988        5.1077        26.0890        0.2902
```

> summary(res)

```
--------------------------------------------
Tobit 2 model with interval outcome (sample selection model)
Maximum Likelihood estimation
BHHH maximisation, 21 iterations
Return code 2: successive function values within tolerance limit
Log-Likelihood: -275.395
300 observations (91 censored and 209 observed)
Intervals of the dependent variable of the outcome equation:
         YO lower upper count
1  (-Inf,5]  -Inf     5    26
2    (5,15]     5    15   130
3 (15, Inf]    15   Inf    53
7 free parameters (df = 293)
Probit selection equation:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.9820     0.1085   9.049  < 2e-16 ***
x1            0.9668     0.1491   6.484 3.78e-10 ***
x2           -1.2862     0.1209 -10.637  < 2e-16 ***
Outcome equation:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  10.2403     0.6681  15.328  < 2e-16 ***
x1            2.6598     0.5921   4.492 1.02e-05 ***
   Error terms:
         Estimate Std. Error t value Pr(>|t|)
logSigma  1.63076    0.07474  21.820  < 2e-16 ***
atanhRho  0.29881    0.36188   0.826    0.410
sigma     5.10774    0.38174  13.380  < 2e-16 ***
sigmaSq  26.08901    3.89971   6.690 1.13e-10 ***
rho       0.29022    0.33140   0.876    0.382
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
--------------------------------------------
```

In the following estimation, the starting values are obtained by a two-step estimation of a tobit-2 model, where the dependent variable of the outcome equation is set to the mid points of the intervals:

```
> res2 <- selection( yS ~ x1 + x2, yO ~ x1, data = dat, boundaries = bound,
+    start = "2step" )
> res2

Call:
 selection(selection = yS ~ x1 + x2, outcome = yO ~ x1, data = dat,      start = "2step",

Coefficients:
S:(Intercept)          S:x1          S:x2  O:(Intercept)          O:x1
       0.9820        0.9668       -1.2862        10.2403        2.6598
     logSigma      atanhRho         sigma        sigmaSq           rho
       1.6308        0.2988        5.1077        26.0890        0.2902

> summary( res2 )

--------------------------------------------
Tobit 2 model with interval outcome (sample selection model)
Maximum Likelihood estimation
BHHH maximisation, 21 iterations
Return code 2: successive function values within tolerance limit
Log-Likelihood: -275.395
300 observations (91 censored and 209 observed)
Intervals of the dependent variable of the outcome equation:
         YO lower upper count
1  (-Inf,5]  -Inf     5    26
2    (5,15]     5    15   130
3 (15, Inf]    15   Inf    53
7 free parameters (df = 293)
Probit selection equation:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.9820     0.1085   9.049  < 2e-16 ***
x1            0.9668     0.1491   6.484 3.78e-10 ***
x2           -1.2862     0.1209 -10.637  < 2e-16 ***
Outcome equation:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  10.2403     0.6681  15.328  < 2e-16 ***
x1            2.6598     0.5921   4.492 1.02e-05 ***
   Error terms:
         Estimate Std. Error t value Pr(>|t|)
logSigma  1.63076    0.07474  21.820  < 2e-16 ***
atanhRho  0.29880    0.36188   0.826    0.410
sigma     5.10774    0.38174  13.380  < 2e-16 ***
sigmaSq  26.08900    3.89970   6.690 1.13e-10 ***
rho       0.29021    0.33140   0.876    0.382
```

```
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
--------------------------------------------
```

The following commands compare the starting values and the estimated coefficients:

```
> # compare starting values (small differences)
> cbind( res$start, res2$start, res$start - res2$start )

                  [,1]       [,2]         [,3]
(Intercept)  0.9818072  0.9800827  0.001724574
x1           0.9663185  0.9686013 -0.002282885
x2          -1.2866893 -1.2808419 -0.005847461
(Intercept) 10.3990738 10.3516223  0.047451510
x1           3.7408112  3.7642665 -0.023455273
logSigma     4.2797893  4.2813219 -0.001532652
atanhRho     0.2362382  0.2550566 -0.018818413

> # combare estimated coefficients (virtually identical)
> cbind( coef( res ), coef( res2 ), coef( res ) - coef( res2 ) )

                  [,1]       [,2]          [,3]
(Intercept)  0.9820207  0.9820209 -2.186942e-07
x1           0.9667788  0.9667789 -1.484320e-07
x2          -1.2862335 -1.2862333 -2.546365e-07
(Intercept) 10.2402931 10.2403019 -8.764296e-06
x1           2.6597945  2.6597881  6.368136e-06
logSigma     1.6307571  1.6307570  1.082407e-07
atanhRho     0.2988073  0.2988007  6.581781e-06
sigma        5.1077402  5.1077397  5.528652e-07
sigmaSq     26.0890100 26.0890044  5.647783e-06
rho          0.2902207  0.2902147  6.027421e-06
```

8 Example with the ‘Smoke’ dataset

The following command loads the dataset:

```
> data( "Smoke" )
```

The following command creates a vector with the boundaries of the intervals:

```
> bounds <- c( 0, 5, 10, 20, 50, Inf )
```

The following command estimates the model with few explanatory variables:

```
> SmokeRes1 <- selection( smoker ~ educ + age,
+   cigs_intervals ~ educ, data = Smoke, boundaries = bounds )
```

The following command estimates the model with more explanatory variables:

```
> SmokeRes2 <- selection( smoker ~ educ + age + restaurn,
+   cigs_intervals ~ educ + income + restaurn, data = Smoke,
+   boundaries = bounds )
```

The following commands test whether adding further explanatory variables significantly improves the explanatory power of the model:

```
> library( "lmtest" )
> lrtest( SmokeRes1, SmokeRes2 )

Likelihood ratio test

Model 1: selection(selection = smoker ~ educ + age, outcome = cigs_intervals ~
    educ, data = Smoke, boundaries = bounds)
Model 2: selection(selection = smoker ~ educ + age + restaurn, outcome = cigs_intervals ~
    educ + income + restaurn, data = Smoke, boundaries = bounds)
  #Df  LogLik Df  Chisq Pr(>Chisq)
1   7 -940.54
2  10 -936.30  3 8.4705    0.03723 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> waldtest( SmokeRes1, SmokeRes2 )

Wald test

Model 1: selection(selection = smoker ~ educ + age, outcome = cigs_intervals ~
    educ, data = Smoke, boundaries = bounds)
Model 2: selection(selection = smoker ~ educ + age + restaurn, outcome = cigs_intervals ~
    educ + income + restaurn, data = Smoke, boundaries = bounds)
  Res.Df Df  Chisq Pr(>Chisq)
1    800
2    797  3 7.8636    0.04892 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Both tests indicate that—at 5% significance level—the model with more explanatory variables (`SmokeRes2`) has significantly higher explanatory power than the model with fewer explanatory variables (`SmokeRes1`).

References

Petersen, S., Henningsen, G., and Henningsen, A. (2017). Can changes in households' life situations predict participation in energy audits and investments in energy savings? Unpublished Manuscript. Department of Management Engineering, Technical University of Denmark.

Sibuya, M. (1960). Bivariate extreme statistics, i. *Annals of the Institute of Statistical Mathematics*, 11(2):195 – 210.

Sungur, E. (1990). Dependence information in parameterized copulas. *Communications in Statistics - Simulation and Computation*, 19(4):1339 – 1360.