Maximum likelihood estimation of Heckman's sample selection model

Herman J. Bierens

October 2007

1 Heckman's sample selection model

1.1 Introduction

Heckman's sample selection model[1] is based on two latent dependent variables models:

$$\begin{aligned} Y_1^* &= \beta' X + U_1, && (1) \\ Y_2^* &= \gamma' Z + U_{2,} && (2) \end{aligned}$$

where X and Z are vectors of regressors, possibly containing common components, including intercepts, and the errors U_1 and U_2 are, conditional on X and Z, jointly bivariate normally distributed with zero mean vector and variance matrix Σ.

The model for Y_1^* is the one we are interested in, but Y_1^* is only observable if $Y_2^* > 0$. Thus the observed dependent variable Y is

$$\begin{aligned} Y &= Y_1^* \text{ if } Y_2^* > 0, \\ Y &= \text{missing value if } Y_2^* \leq 0. \end{aligned}$$

However, the Z's are observable if Y is a missing value, and the X's are observable if the Y's are.

[1] Heckman, James J. (1979): "Sample Selection Bias as a Specification Error", *Econometrica* 47, 153-161. (Heckman got the Nobel prize for this article.)

The variance matrix Σ can be written as

$$\Sigma = \Delta\Delta',$$

where Δ is an upper-triangular matrix:

$$\Delta = \begin{pmatrix} \delta_1 & \delta_2 \\ 0 & \delta_3 \end{pmatrix}$$

Consequently, we can write

$$\begin{aligned} U_1 &= \delta_1 e_1 + \delta_2 e_2, \\ U_2 &= \delta_3 e_2, \end{aligned}$$

where e_1 and e_2 are independent standard normally distributed. Thus the latent dependent variables models (1) and (2) can be written as

$$\begin{aligned} Y_1^* &= \beta' X + \delta_1 e_1 + \delta_2 e_2, && (3) \\ Y_2^* &= \gamma' Z + \delta_3 e_2. && (4) \end{aligned}$$

Without loss of generality we may assume that $\delta_1 > 0$, and since only the sign of Y_2^* plays a role, we may set $\delta_3 = 1$. Then the conditional probability of a missing value of Y is:

$$\begin{aligned} P[Y_2^* &\leq 0|Z, X] = P[e_2 \leq -\gamma' Z] \\ &= 1 - P[e_2 \geq -\gamma' Z] \\ &= 1 - P[-e_2 \leq \gamma' Z] \\ &= 1 - F\left(\gamma' Z\right), \end{aligned}$$

where F is the distribution function of the standard normal distribution, i.e.,

$$F(x) = \int_{-\infty}^{x} f(u) du, \qquad (5)$$

with

$$f(x) = \frac{\exp\left[-x^2/2\right]}{\sqrt{2\pi}}. \qquad (6)$$

Thus, from now on I will assume that

$$\delta_1 > 0, \ \delta_3 = 1.$$

Let D be a dummy variable taking the value 1 if Y is observed, and 0 if not. Then

$$P\left[D=1|\,Z,X\right]=F\left(\gamma'Z\right). \tag{7}$$

The distribution function of Y **conditional** on the event $D=1$ and X and Z is now given by

$$\begin{aligned}
H(y|X,Z) &= P\left[Y\leq y|\,D=1,X,Z\right] \qquad (8)\\
&= \frac{P\left[Y\leq y \text{ and } D=1|\,X,Z\right]}{P\left[D=1|\,X,Z\right]}\\
&= \frac{P\left[Y_1^*\leq y \text{ and } Y_2^*>0|\,X,Z\right]}{F\left(\gamma'Z\right)}\\
&= \frac{P\left[\delta_2 e_2\leq y-\beta'X-\delta_1 e_1 \text{ and } -\gamma'Z<e_2|\,X,Z\right]}{F\left(\gamma'Z\right)}
\end{aligned}$$

1.2 The case $\delta_2>0$

In order to evaluate expression (8) further, and derive the corresponding conditional density, assume first that $\delta_2>0$. Then (8) times $F\left(\gamma'Z\right)$ becomes

$$\begin{aligned}
F\left(\gamma'Z\right).H(y|X,Z) &= P\left[-\gamma'Z<e_2\leq(y-\beta'X-\delta_1e_1)/\delta_2|\,X,Z\right]\\
&= \int_{-\infty}^{\infty}P\left[-\gamma'Z<e_2\leq(y-\beta'X-\delta_1u)/\delta_2|\,X,Z\right]\\
&\quad\times f(u)du\\
&= \int_{-\infty}^{(y-\beta'X+\delta_2\gamma'Z)/\delta_1}\left[F\left((y-\beta'X-\delta_1u)/\delta_2\right)-F\left(-\gamma'Z\right)\right]\\
&\quad\times f(u)du\\
&= \int_{-\infty}^{(y-\beta'X+\delta_2\gamma'Z)/\delta_1}F\left((y-\beta'X-\delta_1u)/\delta_2\right)f(u)du\\
&\quad -F\left(-\gamma'Z\right)F\left((y-\beta'X+\delta_2\gamma'Z)/\delta_1\right)\\
&= \frac{\delta_2}{\delta_1}\int_{-\gamma'Z}^{\infty}F\left(v\right)f((y-\beta'X-\delta_2v)/\delta_1)dv\\
&\quad -F\left(-\gamma'Z\right)F\left((y-\beta'X+\delta_2\gamma'Z)/\delta_1\right)\\
&= -\int_{-\gamma'Z}^{\infty}F\left(v\right)\frac{\partial F((y-\beta'X-\delta_2v)/\delta_1)}{\partial v}dv\\
&\quad -F\left(-\gamma'Z\right)F\left((y-\beta'X+\delta_2\gamma'Z)/\delta_1\right)
\end{aligned}$$

$$
\begin{aligned}
&= -F(v)\,F((y-\beta' X-\delta_2 v)/\delta_1)|_{-\gamma' Z}^{\infty} \\
&\quad + \int_{-\gamma' Z}^{\infty} F\left((y-\beta' X-\delta_2 v)/\delta_1\right) f(v)dv \\
&\quad - F\left(-\gamma' Z\right) F\left((y-\beta' X+\delta_2 \gamma' Z)/\delta_1\right) \\
&= \int_{-\gamma' Z}^{\infty} F\left((y-\beta' X-\delta_2 v)/\delta_1\right) f(v)dv
\end{aligned}
$$

The fifth equality follows by substituting

$$u = (y-\beta' X-\delta_2 v)/\delta_1,$$

and the last two equalities follow from integration by parts.

The corresponding conditional density is now

$$
\begin{aligned}
h(y|X,Z) &= \frac{\partial H(y|X,Z)}{\partial y} \\
&= \frac{1}{\delta_1 F\left(\gamma' Z\right)} \int_{-\gamma' Z}^{\infty} f\left((y-\beta' X-\delta_2 v)/\delta_1\right) f(v)dv
\end{aligned}
$$

It can be shown (see Appendix 1) that for the standard normal density f,

$$
\begin{aligned}
\int_{c}^{\infty} f(a+b.x)f(x)dx &= \frac{f\left(a/\sqrt{b^2+1}\right)}{\sqrt{b^2+1}} \\
&\quad \times \left[1-F\left(c\sqrt{b^2+1}+ab/\sqrt{b^2+1}\right)\right]
\end{aligned}
\tag{9}
$$

Substituting $c = -\gamma' Z$, $a = (y-\beta' X)/\delta_1$ and $b = -\delta_2/\delta_1$, i.e.,

$$
\begin{aligned}
\frac{1}{\sqrt{b^2+1}} &= \frac{\delta_1}{\sqrt{\delta_1^2+\delta_2^2}}, \\
\frac{a}{\sqrt{b^2+1}} &= \frac{y-\beta' X}{\sqrt{\delta_1^2+\delta_2^2}}, \\
c\sqrt{b^2+1}+\frac{ab}{\sqrt{b^2+1}} &= -\frac{\delta_2(y-\beta' X)+\left(\delta_1^2+\delta_2^2\right)\gamma' Z}{\delta_1\sqrt{\delta_1^2+\delta_2^2}},
\end{aligned}
$$

it follows therefore that

$$
h(y|X,Z) = \frac{f\left((y-\beta' X)/\sqrt{\delta_1^2+\delta_2^2}\right)}{\sqrt{\delta_1^2+\delta_2^2}F\left(\gamma' Z\right)}
\tag{10}
$$

$$
\begin{aligned}
& \times\left(1-F\left(-\frac{\delta_2(y-\beta'X)+(\delta_1^2+\delta_2^2)\gamma'Z}{\delta_1\sqrt{\delta_1^2+\delta_2^2}}\right)\right) \\
= \; & \frac{f\left((y-\beta'X)/\sqrt{\delta_1^2+\delta_2^2}\right)}{\sqrt{\delta_1^2+\delta_2^2}F\left(\gamma'Z\right)} \\
& \times F\left(\frac{\delta_2(y-\beta'X)+(\delta_1^2+\delta_2^2)\gamma'Z}{\delta_1\sqrt{\delta_1^2+\delta_2^2}}\right)
\end{aligned}
$$

1.3 The case $\delta_2 < 0$

If $\delta_2 < 0$ then (8) times $F\left(\gamma'Z\right)$ becomes

$$
\begin{aligned}
F\left(\gamma'Z\right).H(y|X,Z) &= P\left[\delta_2 e_2 \le y-\beta'X-\delta_1 e_1 \text{ and } -\delta_2\gamma'Z > \delta_2 e_2 |\, X,Z\right] \\
&= P\left[\delta_2 e_2 \le \min\left((y-\beta'X-\delta_1 e_1), |\delta_2|\gamma'Z\right)\right] \\
&= P\left[|\delta_2| e_2 \le \min\left((y-\beta'X-\delta_1 e_1), |\delta_2|\gamma'Z\right)\right] \\
&= P\left[e_2 \le \min\left((y-\beta'X-\delta_1 e_1)/|\delta_2|, \gamma'Z\right)\right] \\
&= \int_{-\infty}^{\infty} F\left(\min\left((y-\beta'X-\delta_1 u)/\left|\delta_2\right|, \gamma'Z\right)\right) f(u)du \\
&= F\left(\gamma'Z\right)\int_{-\infty}^{(y-\beta'X-|\delta_2|\gamma'Z)/\delta_1} f(u)du \\
&\quad + \int_{(y-\beta'X-|\delta_2|\gamma'Z)/\delta_1}^{\infty} F\left((y-\beta'X-\delta_1 u)/\left|\delta_2\right|\right) f(u)du \\
&= F\left(\gamma'Z\right) F\left((y-\beta'X-\left|\delta_2\right|\gamma'Z)/\delta_1\right) \\
&\quad + \frac{|\delta_2|}{\delta_1}\int_{-\infty}^{\gamma'Z} F\left(v\right) f((y-\beta'X-\left|\delta_2\right| v)/\delta_1)dv \\
&= F\left(\gamma'Z\right) F\left((y-\beta'X-\left|\delta_2\right|\gamma'Z)/\delta_1\right) \\
&\quad - \int_{-\infty}^{\gamma'Z} F\left(v\right) \frac{\partial F((y-\beta'X-\left|\delta_2\right| v)/\delta_1)}{\partial v} dv \\
&= F\left(\gamma'Z\right) F\left((y-\beta'X-\left|\delta_2\right|\gamma'Z)/\delta_1\right) \\
&\quad - F\left(v\right) F((y-\beta'X-\left|\delta_2\right| v)/\delta_1)\big|_{-\infty}^{\gamma'Z} \\
&\quad + \int_{-\infty}^{\gamma'Z} F((y-\beta'X-\left|\delta_2\right| v)/\delta_1) f\left(v\right) dv \\
&= \int_{-\infty}^{\gamma'Z} F((y-\beta'X-\left|\delta_2\right| v)/\delta_1) f\left(v\right) dv
\end{aligned}
$$

The corresponding conditional density is now

$$\begin{aligned}
h(y|X,Z) &= \frac{\partial H(y|X,Z)}{\partial y} \\
&= \frac{1}{\delta_1 F(\gamma' Z)} \int_{-\infty}^{\gamma' Z} f((y - \beta' X - |\delta_2|\, v)/\delta_1) f(v)\, dv
\end{aligned}$$

It can be shown (see Appendix 1) that for the standard normal density f,

$$\begin{aligned}
\int_{-\infty}^{c} f(a + b.x) f(x) dx &= \frac{f\left(a/\sqrt{b^2+1}\right)}{\sqrt{b^2+1}} \\
&\quad \times F\left(c\sqrt{b^2+1} + ab/\sqrt{b^2+1}\right)
\end{aligned} \tag{11}$$

Substituting $c = \gamma' Z$, $a = (y - \beta' X)/\delta_1$ and $b = -|\delta_2|/\delta_1$, i.e.,

$$\begin{aligned}
\frac{1}{\sqrt{b^2+1}} &= \frac{\delta_1}{\sqrt{\delta_1^2 + \delta_2^2}}, \\
\frac{a}{\sqrt{b^2+1}} &= \frac{y - \beta' X}{\sqrt{\delta_1^2 + \delta_2^2}}, \\
c\sqrt{b^2+1} + \frac{ab}{\sqrt{b^2+1}} &= \frac{\sqrt{\delta_1^2 + \delta_2^2}}{\delta_1}\gamma' Z + \frac{-|\delta_2|\,(y - \beta' X)}{\delta_1 \sqrt{\delta_1^2 + \delta_2^2}} \\
&= \frac{\delta_2 (y - \beta' X) - (\delta_1^2 + \delta_2^2)\, \gamma' Z}{\delta_1 \sqrt{\delta_1^2 + \delta_2^2}},
\end{aligned}$$

it follows that

$$\begin{aligned}
h(y|X,Z) &= \frac{f\left((y - \beta' X)/\sqrt{\delta_1^2 + \delta_2^2}\right)}{\sqrt{\delta_1^2 + \delta_2^2} F(\gamma' Z)} \\
&\quad \times F\left(\frac{\delta_2 (y - \beta' X) + (\delta_1^2 + \delta_2^2) \gamma' Z}{\delta_1 \sqrt{\delta_1^2 + \delta_2^2}}\right)
\end{aligned}$$

which is the same as in the case $\delta_2 > 0$.

1.4 The conditional density of the observed Y

Next, substitute

$$\delta_1 = \sigma\sqrt{1-\rho^2},\ \delta_2 = \rho\sigma,$$

which correspond to

$$\Sigma = \Delta\Delta' = \begin{pmatrix} \delta_1^2 + \delta_2^2 & \delta_2 \\ \delta_2 & 1 \end{pmatrix} = \begin{pmatrix} \sigma^2 & \sigma\rho \\ \sigma\rho & 1 \end{pmatrix},$$

where σ^2 is the variance of U_1 and $\rho \in (-1,1)$ is the correlation between U_1 and U_2. Then (10) simplifies to:

$$h(y|X,Z,\beta,\gamma,\rho,\sigma) = \frac{f\left((y-\beta'X)/\sigma\right)}{\sigma F\left(\gamma'Z\right)}.F\left(\frac{\rho(y-\beta'X)/\sigma + \gamma'Z}{\sqrt{1-\rho^2}}\right). \quad (12)$$

The case $\delta_2 = 0$ corresponds to $\rho = 0$:

$$h(y|X,Z,\beta,\gamma,\rho,\sigma) = f\left((y-\beta'X)/\sigma\right)/\sigma,$$

which is just the conditional density of Y_1^*.

2 Sample selection bias

The conditional expectation corresponding to (12) is

$$E\left[Y|D=1,X,Z\right] = \beta'X + \sigma\rho\frac{f\left(\gamma'Z\right)}{F\left(\gamma'Z\right)} \quad (13)$$

and the conditional variance involved is

$$Var\left[Y|D=1,X,Z\right] = \sigma^2 - \rho^2\sigma^2\left(\gamma'Z + \frac{f\left(\gamma'Z\right)}{F\left(\gamma'Z\right)}\right)\frac{f\left(\gamma'Z\right)}{F\left(\gamma'Z\right)}.. \quad (14)$$

See Appendix 2. Thus

$$E\left[Y|D=1,X\right] = \beta'X + \sigma\rho E\left[f\left(\gamma'Z\right)/F\left(\gamma'Z\right)|\,X\right]. \quad (15)$$

The second term is the cause of the sample selection bias of the OLS estimator of β if Y is regressed on X using the valid observations on Y only.

Note that if X and Z are independent then

$$E\left[f\left(\gamma'Z\right)/F\left(\gamma'Z\right)|\,X\right] = E\left[f\left(\gamma'Z\right)/F\left(\gamma'Z\right)\right]$$

is constant, and therefore only affects the intercept.

3 The log-likelihood function and score vector

Let for $j = 1, ..., n$, $D_j = 1$ if Y_j is observed, and $D_j = 0$ if not. The regressors $X_j \in \mathbb{R}^k$ are observable if the corresponding Y_j are observable, and the $Z_j \in R^\ell$ are observable for all j. It will be assumed that the data involved is a random sample with non-response for Y_j if $D_j = 0$.

Without loss of generality we may assume that $Y_j = 0$ if $D_j = 0$. The actual dependent variable is now the pair $(D_j, D_j Y_j)$, with joint conditional distribution given by

$$\begin{aligned}
&\frac{d}{dy} P\left[D_j = 1, D_j Y_j \leq y | X_j, Z_j\right] \\
&= \frac{d}{dy} P\left[Y_j \leq y | D_j = 1, X_j, Z_j\right] P\left[D_j = 1 | X_j, Z_j\right] \\
&= h(y|X, Z, \beta, \gamma, \sigma, \rho) F\left(\gamma' Z\right)
\end{aligned}$$

and

$$P\left[D_j = 0, D_j Y_j = 0 | X_j, Z_j\right] = P\left[D_j = 0 | X_j, Z_j\right] = 1 - F\left(\gamma' Z\right)$$

Then the log-likelihood takes the form

$$\begin{aligned}
\ln \mathcal{L}(\theta) &= \sum_{j=1}^{n} (1 - D_j) \ln\left(1 - F(\gamma' Z_j)\right) + \sum_{j=1}^{n} D_j \ln\left(F\left(\gamma' Z_j\right)\right) \qquad (16) \\
&\quad + \sum_{j=1}^{n} D_j \ln\left(h(Y_j | X_j, Z_j, \beta, \gamma, \rho, \sigma)\right),
\end{aligned}$$

where

$$\theta = (\beta', \gamma', \sigma, \rho)'. \qquad (17)$$

The corresponding score vector $\partial \ln \mathcal{L}(\theta)/\partial\theta$ is .

$$\frac{\partial \ln \mathcal{L}(\theta)}{\partial \theta'} = \sum_{j=1}^{n} \delta_j(\theta) \qquad (18)$$

where

$$\delta_j(\theta) = \left(D_j \frac{f\left(\gamma' Z_j\right)}{F\left(\gamma' Z_j\right)} - (1 - D_j) \frac{f(\gamma' Z_j)}{1 - F(\gamma' Z_j)}\right) \begin{pmatrix} 0_k \\ Z_j \\ 0 \\ 0 \end{pmatrix} \qquad (19)$$

$$+D_j \frac{\partial \ln \left(h(Y_j|X_j, Z_j, \beta, \gamma, \rho, \sigma)\right)}{\partial \theta'}$$

with 0_k a k-vectorsof zeros. The partial derivative vector in (19) is derived in Appendix 3.

Moreover, recall from maximum likelihood theory that for the true parameter vector θ_0,

$$H = \lim_{n\to\infty} E\left[-\frac{1}{n}\left.\frac{\partial^2 \ln \mathcal{L}(\theta)}{\partial\theta\partial\theta'}\right|_{\theta=\theta_0}\right] = \lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n} E\left[\delta_j(\theta_0)\delta_j(\theta_0)'\right],$$

and that under some regularity conditions the maximum likelihood estimator $\widehat{\theta}$ of θ_0 satisfies

$$\sqrt{n}\left(\widehat{\theta}-\theta_0\right) \to N\left[0, H^{-1}\right]$$

in distribution, where H can be consistently estimated by

$$\widehat{H} = \frac{1}{n}\sum_{j=1}^{n}\delta_j(\widehat{\theta})\delta_j(\widehat{\theta})'.$$

4 Initial parameter estimates

The log-likelihood function (16) is highly nonlinear in the parameters, and even more so are the components of the score vector $\partial \ln \mathcal{L}(\theta)/\partial\theta$ (see Appendix 3) and the elements of the Hessian matrix $\frac{\partial^2 \ln \mathcal{L}(\theta)}{\partial\theta\partial\theta'}$. Moreover, the latter matrix may not be negative definite for all values of θ (at least I could not verify this). Therefore, EasyReg maximizes the log-likelihood function (16) by using the simplex method of Nelder and Mead, which only requires evalutions of (16) itself. However, this method is rather slow, and if the Hessian is not negative definite for all values of θ one may get stuck in a local optimum. Therefore, it is important to start the simplex iteration from a starting value of θ already close to the true parameter value θ_0. Such a starting value $\widetilde{\theta}$, say, can be derived as follows.

The parameter vector γ can be estimated by Probit analysis. Given the Probit estimator $\widetilde{\gamma}$, say, the parameter vector β and the parameter $\alpha = \sigma\rho$ can be estimated by regressing Y_j on X_j and $f\left(\widetilde{\gamma}'Z_j\right)/F\left(\widetilde{\gamma}'Z_j\right)$ for the observations j for which $D_j = 1$, with OLS estimators $\widetilde{\beta}$ and $\widetilde{\alpha}$, and residual $\widetilde{v}_j$:

$$Y_j = \widetilde{\beta}'X_j + \widetilde{\alpha} f\left(\widetilde{\gamma}'Z_j\right)/F\left(\widetilde{\gamma}'Z_j\right) + \widetilde{v}_j.$$

Now (14) suggests to estimate σ^2 by

$$\begin{aligned}
\widetilde{\sigma}^2 &= \frac{1}{m}\sum_{j=1}^{m} D_j \left[\widetilde{v}_j^2 + \widetilde{\alpha}^2 \left(\frac{1}{m}\sum_{j=1}^{m} \widetilde{\gamma}' Z_j \frac{f\left(\widetilde{\gamma}' Z_j\right)}{F\left(\widetilde{\gamma}' Z_j\right)} + \frac{1}{m}\sum_{j=1}^{m} \frac{f\left(\widetilde{\gamma}' Z_j\right)^2}{F\left(\widetilde{\gamma}' Z_j\right)^2}\right)\right] \\
&= \frac{1}{m}\sum_{j=1}^{m} D_j \widetilde{v}_j^2 + \widetilde{\alpha}^2 \left(\frac{1}{m}\sum_{j=1}^{m} D_j \left[\left(\widetilde{\gamma}' Z_j\right) F\left(\widetilde{\gamma}' Z_j\right) + f\left(\widetilde{\gamma}' Z_j\right)\right] \frac{f\left(\widetilde{\gamma}' Z_j\right)}{F\left(\widetilde{\gamma}' Z_j\right)^2}\right).
\end{aligned}$$

where

$$m = \sum_{j=1}^{n} D_j.$$

Note that $\widetilde{\sigma}^2 \geq \widetilde{\alpha}^2$, because

$$\inf_u \left[uF(u) + f(u)\right] = \lim_{u \to -\infty} \left[uF(u) + f(u)\right] = 0.$$

Finally, ρ can be estimated by

$$\widetilde{\rho} = \widetilde{\alpha}/\sqrt{\widetilde{\sigma}^2}.$$

Let $\widetilde{\theta} = (\widetilde{\beta}', \widetilde{\gamma}', \widetilde{\sigma}, \widetilde{\rho})'$. Under some regularity conditions (one of them is that $m/n \to \lambda \in (0,1)$ as $n \to \infty$) it can be shown that

$$\widetilde{\theta} - \theta_0 = O_p\left(1/\sqrt{n}\right),$$

where θ_0 is the true parameter vector.

5 Appendix 1: Products of normal densities

Let $f(x)$ be the standard normal density. Then

$$\begin{aligned}
f(a + b.x)f(x) &= \frac{\exp\left[-\frac{1}{2}(a + bx)^2 - \frac{1}{2}x^2\right]}{2\pi} \\
&= \frac{\exp\left[-\frac{1}{2}\left(a^2 + 2abx + (1 + b^2)x^2\right)\right]}{2\pi} \\
&= \frac{\exp\left[-\frac{1}{2}\left(\frac{a^2}{1+b^2} + 2\frac{ab}{1+b^2}x + x^2\right)(1 + b^2)\right]}{2\pi}
\end{aligned}$$

$$= \frac{\exp\left[-\frac{1}{2}\left(\left(\frac{ab}{1+b^2}\right)^2 + 2\frac{ab}{1+b^2}x + x^2\right)(1+b^2)\right]}{2\pi}$$

$$\times \exp\left[-\frac{1}{2}\left(\frac{a^2}{1+b^2} - \left(\frac{ab}{1+b^2}\right)^2\right)(1+b^2)\right]$$

$$= \frac{\exp\left[-\frac{1}{2}\left(x+\frac{ab}{1+b^2}\right)^2 / \frac{1}{1+b^2}\right]}{\frac{1}{\sqrt{1+b^2}}\sqrt{2\pi}}$$

$$\times \frac{\exp\left[-\frac{1}{2}\left(\frac{a^2}{1+b^2}\right)\right]}{\sqrt{1+b^2}\sqrt{2\pi}}$$

$$= f\left(\left(x+\frac{ab}{1+b^2}\right)/\frac{1}{\sqrt{1+b^2}}\right) f\left(a/\sqrt{1+b^2}\right)$$

Hence:

$$\int_{-\infty}^{c} f(a+b.x)f(x)dx \tag{20}$$

$$= \int_{-\infty}^{c} f\left(\left(x+\frac{ab}{1+b^2}\right)/\frac{1}{\sqrt{1+b^2}}\right) dx \times f\left(a/\sqrt{1+b^2}\right)$$

$$= \int_{-\infty}^{c} f\left(\left(x+\frac{ab}{1+b^2}\right)/\frac{1}{\sqrt{1+b^2}}\right) d\left(x+\frac{ab}{1+b^2}\right)$$

$$\times f\left(a/\sqrt{1+b^2}\right)$$

$$= \int_{-\infty}^{c+ab/(1+b^2)} f\left(u\sqrt{1+b^2}\right) du$$

$$\times f\left(a/\sqrt{1+b^2}\right)$$

$$= \int_{-\infty}^{c+ab/(1+b^2)} f\left(u\sqrt{1+b^2}\right) d\left(u\sqrt{1+b^2}\right) \times \frac{f\left(a/\sqrt{1+b^2}\right)}{\sqrt{1+b^2}}$$

$$= F\left(c\sqrt{1+b^2}+\frac{ab}{\sqrt{1+b^2}}\right)\frac{f\left(a/\sqrt{1+b^2}\right)}{\sqrt{1+b^2}}$$

This result proves (11).

Setting $c=\infty$ in (20) it follows that

$$\int_{-\infty}^{\infty} f(a+b.x)f(x)dx = \frac{f\left(a/\sqrt{1+b^2}\right)}{\sqrt{1+b^2}}, \tag{21}$$

hence

$$\begin{aligned}
&\int_c^\infty f(a+b.x)f(x)dx \\
&= \int_{-\infty}^\infty f(a+b.x)f(x)dx - \int_{-\infty}^c f(a+b.x)f(x)dx \\
&= \left[1 - F\left(c\sqrt{1+b^2} + \frac{ab}{\sqrt{1+b^2}}\right)\right] \frac{f\left(a/\sqrt{1+b^2}\right)}{\sqrt{1+b^2}}.
\end{aligned}$$

This result proves (9).

6 Appendix 2: The conditional moment generating function and its derivatives

In order to derive the conditional expectation $E[Y|D=1,X,Z]$ and the conditional variance $Var[Y|D=1,X,Z]$ we now compute the moment generating function of the conditional density $h(y|X,Z,\beta,\gamma,\rho,\sigma)$:

$$\begin{aligned}
&m(\xi|X,Z,\beta,\gamma,\rho,\sigma) \\
&= \int_{-\infty}^\infty \exp(\xi y)h(y|X,Z,\beta,\gamma,\rho,\sigma)dy \\
&= \int_{-\infty}^\infty \exp(\xi y)\frac{f\left((y-\beta'X)/\sigma\right)}{\sigma F\left(\gamma'Z\right)} F\left(\frac{\rho(y-\beta'X)/\sigma + \gamma'Z}{\sqrt{1-\rho^2}}\right) dy \\
&= \frac{\exp(\xi\beta'X)}{F\left(\gamma'Z\right)} \int_{-\infty}^\infty \exp(\xi\sigma u) f\left(u\right) F\left(\frac{\rho u + \gamma'Z}{\sqrt{1-\rho^2}}\right) du \\
&= \frac{\exp(\xi\beta'X + \xi^2\sigma^2/2)}{F\left(\gamma'Z\right)} \int_{-\infty}^\infty f\left(u - \xi\sigma\right) F\left(\frac{\rho u + \gamma'Z}{\sqrt{1-\rho^2}}\right) du \\
&= \frac{\exp(\xi\beta'X + \xi^2\sigma^2/2)}{F\left(\gamma'Z\right)} \int_{-\infty}^\infty f\left(u\right) F\left(\frac{\rho u + \rho\xi\sigma + \gamma'Z}{\sqrt{1-\rho^2}}\right) du
\end{aligned}$$

The fourth equality follows from

$$\exp(\xi\sigma u)f\left(u\right) \quad = \quad \frac{\exp\left[-\left(\xi^2\sigma^2 - 2\xi\sigma u + u^2\right)/2\right]}{\sqrt{2\pi}}$$

$$
\begin{aligned}
& \times \exp\left(\xi^2\sigma^2/2\right) \\
= \quad & \exp\left(\xi^2\sigma^2/2\right) f(u-\xi\sigma)
\end{aligned}
$$

Thus,

$$
\begin{aligned}
& \frac{\partial m(\xi|X,Z,\beta,\gamma,\rho,\sigma)}{\partial \xi} && (22) \\
& = (\beta' X + \xi\sigma^2)\frac{\exp(\xi\beta' X + \xi^2\sigma^2/2)}{F\left(\gamma' Z\right)} \int_{-\infty}^{\infty} f\left(u\right) F\left(\frac{\rho u + \rho\xi\sigma + \gamma' Z}{\sqrt{1-\rho^2}}\right) du \\
& + \frac{\rho\sigma \exp(\xi\beta' X + \xi^2\sigma^2/2)}{F\left(\gamma' Z\right)\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} f\left(u\right) f\left(\frac{\rho u + \rho\xi\sigma + \gamma' Z}{\sqrt{1-\rho^2}}\right) du \\
& = (\beta' X + \xi\sigma^2) m(\xi|X,Z,\beta,\gamma,\rho,\sigma) \\
& + \rho\sigma \exp(\xi\beta' X + \xi^2\sigma^2/2)\frac{f\left(\rho\xi\sigma + \gamma' Z\right)}{F\left(\gamma' Z\right)}.
\end{aligned}
$$

The last equality follows from (20) with $c = \infty$, $a = (\rho\xi\sigma + \gamma' Z)/\sqrt{1-\rho^2}$, $b = \rho/\sqrt{1-\rho^2}$:

$$
\int_{-\infty}^{\infty} f\left(u\right) f\left(\frac{\rho u + \rho\xi\sigma + \gamma' Z}{\sqrt{1-\rho^2}}\right) du = \sqrt{1-\rho^2} f\left(\rho\xi\sigma + \gamma' Z\right).
$$

Moreover, it follows from (22) and the easy equality $f'(u) = -uf(u)$ that

$$
\begin{aligned}
& \frac{\partial^2 m(\xi|X,Z,\beta,\gamma,\rho,\sigma)}{(\partial \xi)^2} && (23) \\
& = \sigma^2 m(\xi|X,Z,\beta,\gamma,\rho,\sigma) \\
& + (\beta' X + \xi\sigma^2)\frac{\partial m(\xi|X,Z,\beta,\gamma,\rho,\sigma)}{\partial \xi} \\
& + \rho\sigma(\beta' X + \xi\sigma^2) \exp(\xi\beta' X + \xi^2\sigma^2/2)\frac{f\left(\rho\xi\sigma + \gamma' Z\right)}{F\left(\gamma' Z\right)} \\
& - \rho^2\sigma^2\left(\rho\xi\sigma + \gamma' Z\right) \exp(\xi\beta' X + \xi^2\sigma^2/2)\frac{f\left(\rho\xi\sigma + \gamma' Z\right)}{F\left(\gamma' Z\right)}.
\end{aligned}
$$

Hence

$$
E\left[Y|D=1,X,Z\right] \quad = \quad \left.\frac{\partial m(\xi|X,Z,\beta,\gamma,\rho,\sigma)}{\partial \xi}\right|_{\xi=0}
$$

$$= \quad \beta' X + \rho\sigma \frac{f(\gamma' Z)}{F(\gamma' Z)},$$

$$E\left[Y^2|D=1,X,Z\right] = \left.\frac{\partial^2 m(\xi|X,Z,\beta,\gamma,\rho,\sigma)}{(\partial\xi)^2}\right|_{\xi=0}$$
$$= \sigma^2 + \left(\beta' X + \rho\sigma \frac{f(\gamma' Z)}{F(\gamma' Z)}\right)\beta' X$$
$$+\rho\sigma.\beta' X \frac{f(\gamma' Z)}{F(\gamma' Z)} - \rho^2\sigma^2\gamma' Z \frac{f(\gamma' Z)}{F(\gamma' Z)}$$
$$= \sigma^2 + (\beta' X)^2 + 2\rho\sigma.\beta' X \frac{f(\gamma' Z)}{F(\gamma' Z)} - \rho^2\sigma^2(\gamma' Z)\frac{f(\gamma' Z)}{F(\gamma' Z)}$$
$$= \sigma^2 + \left(\beta' X + \rho\sigma \frac{f(\gamma' Z)}{F(\gamma' Z)}\right)^2$$
$$-\rho^2\sigma^2(\gamma' Z)\frac{f(\gamma' Z)}{F(\gamma' Z)} - \rho^2\sigma^2 \frac{f(\gamma' Z)^2}{F(\gamma' Z)^2},$$

and thus

$$Var\left[Y|D=1,X,Z\right]$$
$$= E\left[Y^2|D=1,X,Z\right] - \left(E\left[Y|D=1,X,Z\right]\right)^2$$
$$= \sigma^2 - \rho^2\sigma^2(\gamma' Z)\frac{f(\gamma' Z)}{F(\gamma' Z)} - \rho^2\sigma^2 \frac{f(\gamma' Z)^2}{F(\gamma' Z)^2}$$

7 Appendix 3: The score vector

In order to derive the score vector $\frac{\partial \ln \mathcal{L}(\theta)}{\partial \theta'}$, I will derive the first-order partial derivatives of

$$\ln h(y|X,Z,\beta,\gamma,\rho,\sigma) \tag{24}$$
$$= \ln f((y-\beta' X)/\sigma) + \ln F\left(\frac{\rho(y-\beta' X)/\sigma + \gamma' Z}{\sqrt{1-\rho^2}}\right)$$
$$-\ln F(\gamma' Z) - \ln\sigma.$$

Using the easy equality $f'(u) = -uf(u)$, it follows from (24) that

$$
\begin{aligned}
&\frac{\partial \ln h(y|X,Z,\beta,\gamma,\rho,\sigma)}{\partial \beta'} \\
&= -\left((y-\beta'X)/\sigma\right)\frac{\partial (y-\beta'X)/\sigma}{\partial \beta'} \\
&+\frac{f\left(\frac{\rho(y-\beta'X)/\sigma+\gamma'Z}{\sqrt{1-\rho^2}}\right)}{F\left(\frac{\rho(y-\beta'X)/\sigma+\gamma'Z}{\sqrt{1-\rho^2}}\right)}\frac{\partial}{\partial \beta'}\left(\frac{\rho(y-\beta'X)/\sigma+\gamma'Z}{\sqrt{1-\rho^2}}\right) \\
&= \frac{1}{\sigma}\left(\frac{y-\beta'X}{\sigma}-\frac{f\left(\frac{\rho(y-\beta'X)/\sigma+\gamma'Z}{\sqrt{1-\rho^2}}\right)}{F\left(\frac{\rho(y-\beta'X)/\sigma+\gamma'Z}{\sqrt{1-\rho^2}}\right)}\frac{\rho}{\sqrt{1-\rho^2}}\right)X \\
&= \frac{1}{\sigma}\left(\frac{y-\beta'X}{\sigma}-g\left(\frac{\rho(y-\beta'X)/\sigma+\gamma'Z}{\sqrt{1-\rho^2}}\right)\frac{\rho}{\sqrt{1-\rho^2}}\right)X\ ,
\end{aligned}
$$

where

$$
g(u)=\frac{f(u)}{F(u)}. \tag{25}
$$

Moreover using the notation (25) it is easy to verify from (24) that

$$
\begin{aligned}
&\frac{\partial \ln h(y|X,Z,\beta,\gamma,\rho,\sigma)}{\partial \gamma'} \\
&= \left[\frac{1}{\sqrt{1-\rho^2}}g\left(\frac{\rho(y-\beta'X)/\sigma+\gamma'Z}{\sqrt{1-\rho^2}}\right)-g\left(\gamma'Z\right)\right]Z,
\end{aligned}
$$

$$
\begin{aligned}
&\frac{\partial \ln h(y|X,Z,\beta,\gamma,\rho,\sigma)}{\partial \rho} \\
&= g\left(\frac{\rho(y-\beta'X)/\sigma+\gamma'Z}{\sqrt{1-\rho^2}}\right)\frac{\partial}{\partial \rho}\left(\frac{\rho(y-\beta'X)/\sigma+\gamma'Z}{\sqrt{1-\rho^2}}\right) \\
&= g\left(\frac{\rho(y-\beta'X)/\sigma+\gamma'Z}{\sqrt{1-\rho^2}}\right)\left(\frac{(y-\beta'X)/\sigma}{\sqrt{1-\rho^2}}\right)
\end{aligned}
$$

$$+g\left(\frac{\rho(y-\beta'X)/\sigma+\gamma'Z}{\sqrt{1-\rho^2}}\right)\frac{\rho\left(\rho(y-\beta'X)/\sigma+\gamma'Z\right)}{(1-\rho^2)\sqrt{1-\rho^2}}$$
$$=g\left(\frac{\rho(y-\beta'X)/\sigma+\gamma'Z}{\sqrt{1-\rho^2}}\right)\frac{(y-\beta'X)/\sigma+\rho\gamma'Z}{(1-\rho^2)\sqrt{1-\rho^2}}$$

and

$$\frac{\partial\ln h(y|X,Z,\beta,\gamma,\rho,\sigma)}{\partial\sigma}$$
$$=-\left((y-\beta'X)/\sigma\right)\frac{\partial\left((y-\beta'X)/\sigma\right)}{\partial\sigma}$$
$$+g\left(\frac{\rho(y-\beta'X)/\sigma+\gamma'Z}{\sqrt{1-\rho^2}}\right)\frac{\partial}{\partial\sigma}\left(\frac{\rho(y-\beta'X)/\sigma+\gamma'Z}{\sqrt{1-\rho^2}}\right)-\frac{1}{\sigma}$$
$$=\frac{1}{\sigma}\left[\left((y-\beta'X)/\sigma\right)^2-1-\frac{\rho}{\sqrt{1-\rho^2}}\left((y-\beta'X)/\sigma\right)\right.$$
$$\left.\times\, g\left(\frac{\rho(y-\beta'X)/\sigma+\gamma'Z}{\sqrt{1-\rho^2}}\right)\right]$$

Given these partial derivatives, the results (18) and (19) follow.