Efficiency Gains via Re-parameterization: The Ordered Probit Model

(K Ch. 9, KPT Ch. 14)

Matlab scripts: `mod11_op_data, mod11_op, mod11_op_paddlers, mod11_op_paddlers_predict`

Matlab functions: `gs_OP`

This model has several interesting features:

- It requires working with limited dependent variables and *latent data*
- It shows how to improve sampler efficiency through *re-parameterization*
- It uses *Metropolis-Hastings* to approximate some conditional posteriors

The OPM often makes sense when we observe integer data that have ordinal, but not cardinal meaning, and we can think of a plausible underlying "latent" continuous construct that drives these scores. For example, in a consumer survey, you might ask buyers to rate a product as "very good, good, average, bad, and very bad". You can easily assign scores from, say, 5 to 1 to these verbal rankings. These scores have ordinal meaning – 4 is higher (better) than 2 – but no cardinal interpretation (We can't say 4 instills exactly twice as much happiness or utility as 2, etc). The latent construct could be "utility" for this example. We start by writing the model as a basic regression in terms of the latent construct:

$$y_i^* = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i \qquad \varepsilon_i \sim n(0,1) \tag{1}$$

We have to set the error variance to 1, as we won't be able to identify it. Next, we specify a relationship between the latent construct y_i^* and the observed score y_i. Assume we have a total of e =1…E observable categories. We translate "observing category e" as y_i^* falling between two thresholds on the real line. Since not all E-1 thresholds are identified, we follow standard convention by setting the first threshold to zero. This yields the following probability brackets:

$$\begin{aligned}
&pr(y_i = 1) = pr(-\infty < y_i^* \le 0) = pr(-\infty < \varepsilon_i \le -\mathbf{x}_i'\boldsymbol{\beta}) = \Phi(-\mathbf{x}_i'\boldsymbol{\beta}) \\
&pr(y_{ip} = 2) = pr(0 < y_i^* \le c_1) = \Phi(c_1 - \mathbf{x}_i'\boldsymbol{\beta}) - \Phi(-\mathbf{x}_i'\boldsymbol{\beta}) \\
&pr(y_{ip} = 3) = pr(c_1 < y_i^* \le c_2) = \Phi(c_2 - \mathbf{x}_i'\boldsymbol{\beta}) - \Phi(c_1 - \mathbf{x}_i'\boldsymbol{\beta}) \\
&\vdots \\
&pr(y_i = E-1) = pr(c_{E-3} < y_i^* \le c_{E-2}) = \Phi(c_{E-2} - \mathbf{x}_i'\boldsymbol{\beta}) - \Phi(c_{E-3} - \mathbf{x}_i'\boldsymbol{\beta}) \\
&pr(y_i = E) = pr(c_{E-2} < y_i^* \le \infty) = 1 - \Phi(c_{E-2} - \mathbf{x}_i'\boldsymbol{\beta}),
\end{aligned} \tag{2}$$

where symbol Φ denotes the standard normal cumulative distribution function. The *c*-terms in (2) are bin-thresholds that are estimated along with the other model parameters. We will index the identified thresholds as $c_b, b = 1\ldots B,$ where $B=E-2$.

The likelihood function for the full model can be written as

$$p(\mathbf{y}\,|\,\boldsymbol{\beta},\mathbf{c}) = \prod_{i=1}^{n}\left[\sum_{e=1}^{E} pr(y_i = e) I(y_i = e)\right] =$$

$$\prod_{i=1}^{n}\left[\begin{array}{l}\prod_{y_i=1}^{E}\Phi(-\mathbf{x}_\mathbf{i}'\boldsymbol{\beta})^{\max(0,1-abs(2(y_i-1))}\left(\Phi(c_1-\mathbf{x}_\mathbf{i}'\boldsymbol{\beta})-\Phi(-\mathbf{x}_\mathbf{i}'\boldsymbol{\beta})\right)^{\max(0,1-abs(2(y_i-2))}\cdots \\ \left(1-\Phi(c_B-\mathbf{x}_\mathbf{i}'\boldsymbol{\beta})\right)^{\max(0,1-abs(2(y_i-E))}\end{array}\right] \tag{3}$$

where **c** comprises all E-2 threshold terms, $pr(y_{ip} = e)$ is given in (2) for any permissible e, and $I(.)$ is an indicator function. Indicator functions are popular in Bayesian notation. An indicator function simply implies that the expression preceded by it is retained in a function if the indicator condition holds, and dropped from the function otherwise. The second expression in (3) accomplishes the same with a more concrete form. The decision to keep or drop a term is now explicitly captured by the term in the exponent (there are probably other ways to do this, but this one will work).

Normally, we would now move on to a discussion of data augmentation, if applicable, and the specification of priors. However, for this model a re-parmaterization as suggested in Nandram and Chen (1996) will bring dramatic gains in efficiency (see KPT Ch. 14 for a comparison). Specifically, we re-parameterize the OP model using the inverse of the highest bin threshold as follows:

$$\begin{array}{ll}\tilde{y}_i^* = \mathbf{x}_\mathbf{i}'\tilde{\boldsymbol{\beta}} + \tilde{\varepsilon}_i & \tilde{\varepsilon}_i \sim n(0,\delta^2), \quad \text{where} \\ \tilde{\boldsymbol{\beta}} = \delta\boldsymbol{\beta}, \quad \tilde{\varepsilon}_i = \delta\varepsilon_i, \quad \text{and } \delta = \dfrac{1}{c_B} & \end{array} \tag{4}$$

This implies the following re-parameterized values for the bin-thresholds:

$$\tilde{c}_b = \delta c_b, b = 1\cdots B-1, \quad \text{and } \tilde{c}_B = 1. \tag{5}$$

The Bayesian estimation program is specified in terms of these transformed parameters. However, the actual posterior parameter values *collected* during simulation are the original terms, i.e. coefficient vector $\boldsymbol{\beta}$ and threshold parameters **c**.

Priors

It is quite popular to specify improper (uniform) priors for the transformed threshold parameters. However, this would restrict our ability to compare models via marginal LHs and Bayes Factors. Thus, we'll choose a reasonably diffuse normal prior for $\tilde{\mathbf{c}}$, truncated to the [0,1] interval (since the lowest threshold is 0 by definition, and the largest threshold is 1 after transformation:

$$\begin{array}{l}\tilde{\boldsymbol{\beta}} \sim mvn(\boldsymbol{\mu}_0,\mathbf{V}_0) \\ \delta^2 \sim ig(v_0,\tau_0), \\ \tilde{c}_j \sim n(0,V_c) I(\tilde{c}_j \in [0,1])\end{array} \tag{6}$$

Data augmentation

The model with likelihood given in (3) has the following joint posterior kernel:

$$
\begin{aligned}
&p\left(\tilde{\boldsymbol{\beta}},\tilde{\mathbf{c}},\delta^2 \mid \mathbf{y},\mathbf{X}\right) \propto \\
&p\left(\tilde{\boldsymbol{\beta}}\right) p\left(\tilde{\mathbf{c}}\right) p\left(\delta^2\right) p\left(\mathbf{y} \mid \tilde{\boldsymbol{\beta}},\tilde{\mathbf{c}},\delta^2,\mathbf{X}\right) = \\
&\exp\left(-\tfrac{1}{2}\left(\tilde{\boldsymbol{\beta}}-\boldsymbol{\mu}_{\mathbf{0}}\right)' \mathbf{V}_{\mathbf{0}}^{-\mathbf{1}}\left(\tilde{\boldsymbol{\beta}}-\boldsymbol{\mu}_{\mathbf{0}}\right)\right) * \left(\delta^2\right)^{-(\nu_0+1)} \exp\left(-\frac{\tau_0}{\sigma^2}\right) * \prod_{b=1}^{B-1} \frac{\exp\left(-\frac{1}{2V_c^2}\tilde{c}_b^2\right)}{F\left(0,V_c,1\right)-F\left(0,V_c,0\right)} * \\
&\prod_{i=1}^{n}\left[\begin{array}{l}\prod_{y_i=1}^{E} \Phi\left(-\mathbf{x}_{\mathbf{i}}'\tilde{\boldsymbol{\beta}}\right)^{\max(0,1-abs(2(y_i-1))}\left(\Phi\left(\tilde{c}_1-\mathbf{x}_{\mathbf{i}}'\tilde{\boldsymbol{\beta}}\right)-\Phi\left(-\mathbf{x}_{\mathbf{i}}'\tilde{\boldsymbol{\beta}}\right)\right)^{\max(0,1-abs(2(y_i-2))}\cdots \\ \left(1-\Phi\left(1-\mathbf{x}_{\mathbf{i}}'\tilde{\boldsymbol{\beta}}\right)\right)^{\max(0,1-abs(2(y_i-E))}\end{array}\right]
\end{aligned}
\tag{7}
$$

where F(.) is the cdf of the normal density with mean 0 and variance V_c, evaluated at 1 and 0, respectively. This would be difficult to analyze via a regular Gibbs Sampler. We will again use the technique of data augmentation, using draws of the transformed *latent dependent variable* $\tilde{\mathbf{y}}^*$:

$$
\begin{aligned}
&p\left(\tilde{\boldsymbol{\beta}},\tilde{\mathbf{c}},\delta^2,\tilde{\mathbf{y}}^* \mid \mathbf{y},\mathbf{X}\right) \propto \\
&p\left(\tilde{\boldsymbol{\beta}}\right) p\left(\tilde{\mathbf{c}}\right) p\left(\delta^2\right) p\left(y^* \mid \tilde{\boldsymbol{\beta}},\delta^2,\mathbf{X}\right) p\left(\mathbf{y} \mid \tilde{\boldsymbol{\beta}},\tilde{\mathbf{c}},\delta^2,\tilde{\mathbf{y}}^*,\mathbf{X}\right)
\end{aligned}
\tag{8}
$$

where

$$
p\left(\tilde{\mathbf{y}}^* \mid \tilde{\boldsymbol{\beta}},\delta^2,\mathbf{X}\right) = \left(2\pi\right)^{-n/2}\left(\delta^2\right)^{-n/2} \exp\left(-\tfrac{1}{2\delta^2}\left(\left(\tilde{\mathbf{y}}^*-\mathbf{X}\tilde{\boldsymbol{\beta}}\right)'\left(\tilde{\mathbf{y}}^*-\mathbf{X}\tilde{\boldsymbol{\beta}}\right)\right)\right) \tag{9}
$$

yields again a "well-behaved" regression model, analogous to the Probit and Tobit case. Similarly, the conditional likelihood function (the last term in (8)) can be written as

$$
p\left(\mathbf{y} \mid \tilde{\mathbf{c}},\tilde{\mathbf{y}}^*\right) = \prod_{i=1}^{n}\left(\begin{array}{l} I\left(y_i=1\right) I\left(-\infty < \tilde{y}_i^* \le 0\right) + I\left(y_i=2\right) I\left(0 < \tilde{y}_i^* \le \tilde{c}_1\right) + \\ I\left(y_i=3\right) I\left(\tilde{c}_1 < \tilde{y}_i^* \le \tilde{c}_2\right) + \cdots + I\left(y_i=E\right) I\left(1 < \tilde{y}_i^* \le \infty\right)\end{array}\right) \tag{10}
$$

As for the Probit and Tobit, this de-links the likelihood from our main parameters of interest, which facilitates their drawing in the GS.

Gibbs Sampler

Conditional on latent data the posterior density for $\tilde{\boldsymbol{\beta}}$ is the same as for the generic normal regression model, i.e.

$$
\begin{aligned}
&p\left(\tilde{\boldsymbol{\beta}} \mid \delta^2,\tilde{\mathbf{y}}^*,\mathbf{X}\right) \sim n\left(\boldsymbol{\mu}_{\mathbf{1}},\mathbf{V}_{\mathbf{1}}\right) \\
&\text{with} \quad \mathbf{V}_{\mathbf{1}} = \left(\mathbf{V}_{\mathbf{0}}^{-\mathbf{1}} + \tfrac{1}{\delta^2}\mathbf{X}'\mathbf{X}\right)^{-1} \text{ and } \quad \boldsymbol{\mu}_{\mathbf{1}} = \mathbf{V}_{\mathbf{1}}\left(\mathbf{V}_{\mathbf{0}}^{-\mathbf{1}}\boldsymbol{\mu}_{\mathbf{0}} + \tfrac{1}{\delta^2}\mathbf{X}'\tilde{\mathbf{y}}^*\right)
\end{aligned}
\tag{11}
$$

The same holds for draws of δ^2, which de facto takes the role of the variance of the regression error:

$$\delta^2 \mid \tilde{\boldsymbol{\beta}}, \tilde{\mathbf{y}}^*, \mathbf{X} \sim ig\left(v_1, \tau_1\right) \quad \text{with}$$

$$v_1 = \frac{2v_0 + n}{2} \qquad \text{and} \qquad \tau_1 = \frac{2\tau_0 + \left(\tilde{\mathbf{y}}^* - \mathbf{X}\tilde{\boldsymbol{\beta}}\right)' \left(\tilde{\mathbf{y}}^* - \mathbf{X}\tilde{\boldsymbol{\beta}}\right)}{2} \tag{12}$$

To derive the conditional posterior kernel for a given transformed threshold parameter $\tilde{c}_b$ we need to work with the following conditional posterior kernel:

$$p\left(\tilde{c}_b \mid \tilde{\mathbf{y}}^*, \tilde{\boldsymbol{\beta}}, \tilde{\mathbf{c}}_{\mathbf{-b}}, \delta^2, \mathbf{X}\right) \propto \frac{\exp\left(-\frac{1}{2V_c^2}\tilde{c}_b^2\right)}{F\left(0,V_c,1\right) - F\left(0,V_c,0\right)} \prod_{i=1}^{n} \left(I\left(\tilde{c}_{b-1} < \tilde{y}_i^* \le \tilde{c}_b\right) + I\left(\tilde{c}_b < \tilde{y}_i^* \le \tilde{c}_{b+1}\right)\right) \tag{13}$$

This implies that we need to work with all observations in the sample for which $y_i = b+1$ (implying $\left(\tilde{c}_{b-1} < \tilde{y}_i^* \le \tilde{c}_b\right)$) and $y_i = b+2$ (implying $\left(\tilde{c}_b < \tilde{y}_i^* \le \tilde{c}_{b+1}\right)$). Again, this is easier analyzed at the observational level. For the case of $y_i = b+1$, i.e. $\left(\tilde{c}_b < \tilde{y}_i^* \le \tilde{c}_{b+1}\right)$, we have

$$\begin{aligned}
&p\left(\tilde{c}_b \mid \tilde{\boldsymbol{\beta}}, \tilde{\mathbf{c}}_{\mathbf{-b}}, \delta^2, \mathbf{x_i}; \tilde{c}_{b-1} < \tilde{y}_i^* \le \tilde{c}_b\right) \propto \frac{\exp\left(-\frac{1}{2V_c^2}\tilde{c}_b^2\right)}{F\left(0,V_c,1\right) - F\left(0,V_c,0\right)} I\left(\tilde{c}_{b-1} < \tilde{y}_i^* \le \tilde{c}_b\right) = \\
&\frac{\exp\left(-\frac{1}{2V_c^2}\tilde{c}_b^2\right)}{F\left(0,V_c,1\right) - F\left(0,V_c,0\right)} \left(F\left(\mathbf{x_i'}\tilde{\boldsymbol{\beta}}, \delta^2, \tilde{c}_b\right) - F\left(\mathbf{x_i'}\tilde{\boldsymbol{\beta}}, \delta^2, \tilde{c}_{b-1}\right)\right) = \\
&\frac{\exp\left(-\frac{1}{2V_c^2}\tilde{c}_b^2\right)}{F\left(0,V_c,1\right) - F\left(0,V_c,0\right)} \left(\Phi\left(\frac{\tilde{c}_b - \mathbf{x_i'}\tilde{\boldsymbol{\beta}}}{\delta}\right) - \Phi\left(\frac{\tilde{c}_{b-1} - \mathbf{x_i'}\tilde{\boldsymbol{\beta}}}{\delta}\right)\right)
\end{aligned} \tag{14}$$

and for the case $y_i = b+2$, i.e. $\left(\tilde{c}_b < \tilde{y}_i^* \le \tilde{c}_{b+1}\right)$ we have

$$\begin{aligned}
&p\left(\tilde{c}_b \mid \tilde{\boldsymbol{\beta}}, \tilde{\mathbf{c}}_{\mathbf{-b}}, \delta^2, \mathbf{x_i}; \tilde{c}_b < \tilde{y}_i^* \le \tilde{c}_{b+1}\right) \propto \frac{\exp\left(-\frac{1}{2V_c^2}\tilde{c}_b^2\right)}{F\left(0,V_c,1\right) - F\left(0,V_c,0\right)} I\left(\tilde{c}_b < \tilde{y}_i^* \le \tilde{c}_{b+1}\right) = \\
&\frac{\exp\left(-\frac{1}{2V_c^2}\tilde{c}_b^2\right)}{F\left(0,V_c,1\right) - F\left(0,V_c,0\right)} \left(\Phi\left(\frac{\tilde{c}_{b+1} - \mathbf{x_i'}\tilde{\boldsymbol{\beta}}}{\delta}\right) - \Phi\left(\frac{\tilde{c}_b - \mathbf{x_i'}\tilde{\boldsymbol{\beta}}}{\delta}\right)\right)
\end{aligned} \tag{15}$$

In essence, the first component is simply the kernel of a truncated normal density, restricted to the [0,1] interval, and the second component is a difference between two *cdf*-expressions. This conditional kernel cannot be "massaged" to yield a convenient form to draw from. A *Metropolis-Hastings* algorithm is needed to obtain draws from this distribution.

For our current objective to draw $\tilde{c}_b$ a simple random-walk type proposal function works. Thus, we draw a candidate $\tilde{c}_{b,c}$ from a normal distribution with mean $\tilde{c}_b$ (the current draw), and a pre-specified standard deviation s_c. The candidate draw is accepted with probability

$$\alpha\left(\tilde{c}_b,\tilde{c}_{b,c}\right)=\min\left\{\frac{p\left(\tilde{c}_{b,c}\mid\tilde{\mathbf{y}}^*,\tilde{\boldsymbol{\beta}},\tilde{\mathbf{c}}_{\text{-}\mathbf{b}},\delta^2,\mathbf{x}_\mathbf{i}\right)}{p\left(\tilde{c}_b\mid\tilde{\mathbf{y}}^*,\tilde{\boldsymbol{\beta}},\tilde{\mathbf{c}}_{\text{-}\mathbf{b}},\delta^2,\mathbf{x}_\mathbf{i}\right)},1\right\}.$$

For example, for the first transformed threshold $\tilde{c}_1$, we first compute all components that enter the acceptance probability that include the current ("old") draw of $\tilde{c}_1$.

$$\begin{aligned}&\log p\left(\tilde{c}_1^{old}\mid\tilde{\boldsymbol{\beta}},\delta,\tilde{c}_2\right)=\\&\log\left(\frac{\exp\left(-\frac{1}{2V_c^2}\left(\tilde{c}_1^{old}\right)^2\right)}{F\left(0,V_c,1\right)-F\left(0,V_c,0\right)}\right)+\sum_{i:y_i=2}\log\left(\Phi\left(\frac{\tilde{c}_1^{old}-\left(\mathbf{x}_\mathbf{i}'\tilde{\boldsymbol{\beta}}\right)}{\delta}\right)-\Phi\left(\frac{-\left(\mathbf{x}_\mathbf{i}'\tilde{\boldsymbol{\beta}}\right)}{\delta}\right)\right)+\\&\sum_{i:y_i=3}\log\left(\Phi\left(\frac{\tilde{c}_2-\left(\mathbf{x}_\mathbf{i}'\tilde{\boldsymbol{\beta}}\right)}{\delta}\right)-\Phi\left(\frac{\tilde{c}_1^{old}-\left(\mathbf{x}_\mathbf{i}'\tilde{\boldsymbol{\beta}}\right)}{\delta}\right)\right)\end{aligned}\tag{16}$$

The trick here is to quickly identify the observations in your sample that fall into the second and third category (as indexed by the subscripts to the summation sign). This can be accomplished in straightforward fashion via the "find" function – see script `OP`.

We then draw a candidate ("new") $\tilde{c}_1$ from the truncated normal density with mean $\tilde{c}_1^{old}$, standard deviation s_c (our "tuning parameter", chosen at the onset), and truncated to the interval $[0,\tilde{c}_2]$. That way we avoid getting a candidate draw that is smaller than the next smaller or larger than the next larger threshold. Thus, we have

$$q\left(\tilde{c}_1^{new}\mid\tilde{c}_1^{old}\right)=n\left(\tilde{c}_1^{old},s_c^2\right)I\left(\tilde{c}_1^{new}\in\left[0,\tilde{c}_2^{old}\right]\right)\tag{17}$$

We then evaluate the posterior kernel at this new draw, i.e we obtain the value of the density

$$\begin{aligned}&\log p\left(\tilde{c}_1^{new}\mid\tilde{\boldsymbol{\beta}},\delta,\tilde{c}_2\right)=\\&\log\left(\frac{\exp\left(-\frac{1}{2V_c^2}\left(\tilde{c}_1^{new}\right)^2\right)}{F\left(0,V_c,1\right)-F\left(0,V_c,0\right)}\right)+\sum_{i:y_i=2}\log\left(\Phi\left(\frac{\tilde{c}_1^{new}-\left(\mathbf{x}_\mathbf{i}'\tilde{\boldsymbol{\beta}}\right)}{\delta}\right)-\Phi\left(\frac{-\left(\mathbf{x}_\mathbf{i}'\tilde{\boldsymbol{\beta}}\right)}{\delta}\right)\right)+\\&\sum_{i:y_i=3}\log\left(\Phi\left(\frac{\tilde{c}_2-\left(\mathbf{x}_\mathbf{i}'\tilde{\boldsymbol{\beta}}\right)}{\delta}\right)-\Phi\left(\frac{\tilde{c}_1^{new}-\left(\mathbf{x}_\mathbf{i}'\tilde{\boldsymbol{\beta}}\right)}{\delta}\right)\right)\end{aligned}\tag{18}$$

The last step of the MH routine is the computation of the acceptance probability (in log form), given as

$$\log\left(\alpha\left(\tilde{c}_1^{old},\tilde{c}_1^{new}\right)\right)=\log p\left(\tilde{c}_1^{new}\mid\tilde{\boldsymbol{\beta}},\delta,\tilde{c}_2\right)-\log p\left(\tilde{c}_1^{old}\mid\tilde{\boldsymbol{\beta}},\delta,\tilde{c}_2\right)\tag{19}$$

We then accept the new draw if $\log\left(\alpha\left(\tilde{c}_1^{old},\tilde{c}_1^{new}\right)\right)$>log(u), where u is a random draw from the [0,1] uniform density.

We then repeat this MH sub-routine for all (transformed) threshold parameters, as shown in the Matlab script.

Draws of latent data

Formally, the conditional posterior kernel for the latent data is given by

$$p\left(\tilde{\mathbf{y}}^* \mid \tilde{\boldsymbol{\beta}},\delta^2,\tilde{\mathbf{c}},\mathbf{y},\mathbf{X}\right)=\left(2\pi\right)^{-n/2}\left(\delta^2\right)^{-n/2}\exp\left(-\tfrac{1}{2\delta^2}\left(\left(\tilde{\mathbf{y}}^*-\mathbf{X}\tilde{\boldsymbol{\beta}}\right)'\left(\tilde{\mathbf{y}}^*-\mathbf{X}\tilde{\boldsymbol{\beta}}\right)\right)\right)* \\ \prod_{i=1}^{n}\begin{pmatrix} I\left(y_i=1\right)I\left(-\infty<\tilde{y}_i^*\le 0\right)+I\left(y_i=2\right)I\left(0<\tilde{y}_i^*\le\tilde{c}_1\right)+ \\ I\left(y_i=3\right)I\left(\tilde{c}_1<\tilde{y}_i^*\le\tilde{c}_2\right)+\cdots+I\left(y_i=E\right)I\left(1<\tilde{y}_i^*\le\infty\right)\end{pmatrix} \tag{20}$$

Again, we need to examine this observation by observation. For example, for the case $y_i=1$, we get

$$p\left(\tilde{y}_i^* \mid \tilde{\boldsymbol{\beta}},\delta^2,\tilde{\mathbf{c}},y_i=1,\mathbf{x_i}\right)=\left(2\pi\right)^{-1/2}\left(\delta^2\right)^{-1/2}\exp\left(-\tfrac{1}{2\delta^2}\left(\left(\tilde{y}_i^*-\mathbf{x_i}'\tilde{\boldsymbol{\beta}}\right)'\left(\tilde{y}_i^*-\mathbf{x_i}'\tilde{\boldsymbol{\beta}}\right)\right)\right)*I\left(-\infty<\tilde{y}_i^*\le 0\right) \tag{21}$$

which implies that we draw the corresponding $\tilde{y}_i^*$ from a normal density, truncated from above at zero, i.e.

$$\tilde{y}_i^* \mid \tilde{\boldsymbol{\beta}},\delta^2,\tilde{\mathbf{c}},y_i=1,\mathbf{x_i} \sim tn\left(\mathbf{x_i}'\tilde{\boldsymbol{\beta}},\delta,-\infty,0\right) \tag{22}$$

Generally, for the case $y_i=b+1$ we must consider

$$p\left(\tilde{y}_i^* \mid \tilde{\boldsymbol{\beta}},\delta^2,\tilde{\mathbf{c}},y_i=b+1,\mathbf{x_i}\right)=\left(2\pi\right)^{-1/2}\left(\delta^2\right)^{-1/2}\exp\left(-\tfrac{1}{2\delta^2}\left(\left(\tilde{y}_i^*-\mathbf{x_i}'\tilde{\boldsymbol{\beta}}\right)'\left(\tilde{y}_i^*-\mathbf{x_i}'\tilde{\boldsymbol{\beta}}\right)\right)\right)*I\left(\tilde{c}_{b-1}<\tilde{y}_i^*\le\tilde{c}_b\right) \tag{23}$$

leading to

$$\tilde{y}_i^* \mid \tilde{\boldsymbol{\beta}},\delta^2,\tilde{\mathbf{c}},y_i=b+1,\mathbf{x_i} \sim tn\left(\mathbf{x_i}'\tilde{\boldsymbol{\beta}},\delta,\tilde{c}_{b-1},\tilde{c}_b\right) \tag{24}$$

See the Matlab script for details.

Predictive constructs of interest

Let's call the generic posterior predictive construct of interest $g\left(y_p \mid \mathbf{y},\mathbf{x_p}\right)$, where $\mathbf{x_p}$ are specific settings for the explanatory variables in the OP model. As we know, the PPD can then be written as

$$p\left(g\left(y_p \mid \mathbf{y},\mathbf{x_p}\right)\right)=\int_{\boldsymbol{\theta}} g\left(y_p \mid \mathbf{x_p},\boldsymbol{\theta}\right)p\left(\boldsymbol{\theta} \mid \mathbf{y},\mathbf{X}\right)d\boldsymbol{\theta} \tag{25}$$

We can now focus on different explicit expressions for $g\left(y_p \mid \mathbf{x_p}, \boldsymbol{\theta}\right)$. In some applications, one might be interested in the probability that an individual's score falls into a certain category, e.g.

$$pr\left(y_p = e \mid \mathbf{x_p}, \boldsymbol{\beta}, \mathbf{c}\right) = pr\left(c_{e-2} < y_p^* \le c_{e-1} \mid \mathbf{x_p}, \boldsymbol{\beta}\right) = \Phi\left(c_{e-1} - \mathbf{x_p'}\boldsymbol{\beta}\right) - \Phi\left(c_{e-2} - \mathbf{x_p'}\boldsymbol{\beta}\right) \tag{26}$$

For example, the probability of a rating of "4" on a 5-point Likert scale is given as

$$pr\left(y_p = 4 \mid \mathbf{x_p}, \boldsymbol{\beta}, \mathbf{c}\right) = pr\left(c_2 < y_p^* \le c_3 \mid \mathbf{x_p}, \boldsymbol{\beta}\right) = \Phi\left(c_3 - \mathbf{x_p'}\boldsymbol{\beta}\right) - \Phi\left(c_2 - \mathbf{x_p'}\boldsymbol{\beta}\right) \tag{27}$$

Similarly, the probability of a rating of "*e*" or higher can be expressed as

$$pr\left(y_p \ge e \mid \mathbf{x_p}, \boldsymbol{\beta}, \mathbf{c}\right) = pr\left(y_p^* > c_{e-2} \mid \mathbf{x_p}, \boldsymbol{\beta}\right) = 1 - \Phi\left(c_{e-2} - \mathbf{x_p'}\boldsymbol{\beta}\right) \tag{28}$$

For example, the probability of a rating of "3" or higher on a 5-point Likert scale is given as

$$pr\left(y_p \ge 3 \mid \mathbf{x_p}, \boldsymbol{\beta}, \mathbf{c}\right) = pr\left(y_p^* > c_1 \mid \mathbf{x_p}, \boldsymbol{\beta}\right) = 1 - \Phi\left(c_1 - \mathbf{x_p'}\boldsymbol{\beta}\right) \tag{29}$$

In some (rare) cases it might make sense to consider the "expected rating" for a person / scenario combination with features $\mathbf{x_p}$. This expectation can be formally expressed as

$$E\left(y_p \mid \mathbf{x_p}, \boldsymbol{\beta}, \mathbf{c}\right) = \sum_{e=E_{\min}}^{E_{\max}} e * pr\left(y_p = e \mid \mathbf{x_p}, \boldsymbol{\beta}, \mathbf{c}\right) \tag{30}$$

where the expression for the last term is given in (26).

Marginal effects

The expressions for the PPD of marginal effects is similar to the one shown for the Probit model. Specifically, for a continuous regressor x_j we have (e.g. Greene, p. 833)

$$\frac{\partial \Pr\left(y_p = e \mid \mathbf{y}, \mathbf{x_p}, \boldsymbol{\beta}, \mathbf{c}\right)}{\partial x_j} = \frac{\partial \Pr\left(-c_{e-2} < y_p^* \le c_{e-1} \mid \mathbf{x_p}, \boldsymbol{\beta}, \mathbf{c}\right)}{\partial x_j} = -\beta_j\left(\phi\left(c_{e-1} - \mathbf{x_p'}\boldsymbol{\beta}\right) - \phi\left(c_{e-2} - \mathbf{x_p'}\boldsymbol{\beta}\right)\right) \tag{31}$$

For binary regressors, it will make more sense to use

$p\left(\Pr\left(y_p = e \mid \mathbf{x_{p,-j}}, x_j = 1\right) - \Pr\left(y_p = e \mid \mathbf{x_{p,-j}}, x_j = 0\right)\right)$ as in the Probit and Tobit models.

Note that for the continuous version in (31) the sign of β_j translates unambiguously into the sign of the marginal effect only for the lowest and highest category, for which we have

$$\frac{\partial \Pr\left(y_p = e_{\min} \mid \mathbf{y}, \mathbf{x_p}, \boldsymbol{\beta}, \mathbf{c}\right)}{\partial x_j} = \frac{\partial \Pr\left(-\infty < y_p^* \le 0 \mid \mathbf{x_p}, \boldsymbol{\beta}, \mathbf{c}\right)}{\partial x_j} = -\beta_j \left(\phi\left(-\mathbf{x}_\mathbf{p}'\boldsymbol{\beta}\right)\right)$$
$$\frac{\partial \Pr\left(y_p = E \mid \mathbf{y}, \mathbf{x_p}, \boldsymbol{\beta}, \mathbf{c}\right)}{\partial x_j} = \frac{\partial \Pr\left(c_{E-2} < y_p^* < \infty \mid \mathbf{x_p}, \boldsymbol{\beta}, \mathbf{c}\right)}{\partial x_j} = \beta_j \left(\phi\left(c_{E-2} - \mathbf{x}_\mathbf{p}'\boldsymbol{\beta}\right)\right) \tag{32}$$

Matlab scripts `mod11_op_paddlers` and `mod11_op_predict` provide examples.

Extension: Hierarchical Ordered Probit Model

Matlab scripts: `mod11_HOP_Pac, mod11_HOP_Mag, mod11_HOP_Car, mod11_HOP_Pac_postest, mod11_HOP_Mag_postest, mod11_HOP_Car_postest`
Matlab functions: `gs_HOP_normal`

A logical extension of the ordered probit model is a variant where some or all of the coefficients in $\boldsymbol{\beta}$ are random, i.e. have a hierarchical distribution just like in the hierarchical regression model discussed previously.

The model is then augmented with draws of the individual $\boldsymbol{\beta}_i$'s in addition to the latent data. Details for the hierarchical ordered probit (HOP) (along with similar models to analyze limited integer data) are given in the working paper by Moeltner et al. (2008), sent to you via e-mail.

The application in that paper and captured in the Matlab files posted on our web site is with respect to a common resource field experiment in Colombia.