# Transformations and priors that we use

January 25, 2014

In this document, we lay out three aspectsof model implementation:

1. Order that parameters are passed in to the estimation algorithm;
2. Transformations that we use to get from unconstrained to constrained parameters; and
3. Prior distributions of the *unconstrained* parameters, including any Jacobians from the transformations.

Parameters are passed in the following order:

1. $\theta_{12}$, columnwise, with no transformation.
2. $\bar{c}$,the average $c_j$, across $J$ brands (no transformation);
3. $\log \operatorname{sd} c$, the log of the standard deviation of $c_j$;
4. $c_{\text{off}}$, the standardized offset of $c_j$ from the mean. Thus, $c_j = \operatorname{sd}(c)\,(c_{\text{off}} - \bar{c})$.
5. $\bar{u}$,the average $u_j$, across $J$ brands (no transformation);
6. $\log \operatorname{sd} u$, the log of the standard deviation of $u_j$;
7. $u_{\text{off}}$, the standardized offset of $u_j$ from the mean. Thus, $u_j = \operatorname{sd}(u)\,(u_{\text{off}} - \bar{u})$.
8. $\phi$, which is included only if the $H$ matrix is included. Columnwise, no transformation;
9. $\operatorname{logit} \delta$, a scalar parameter;
10. log diagonal elements for $V_1$ (see below for details);
11. factors for $V_1$, columnwise, some elements transformed (see details);
12. log diagonal elements for $V_2$ (see below for details);
13. factors for $V_2$, columnwise, some elements transformed (see details);
14. log scale parameter for $W_1$ (see below);
15. transformed Cholesky factors for $W_1$ (see below);
16. log diagonal elements for $W_1$ (see below for details);
17. factors for $W_1$, columnwise, some elements transformed (see details);

## 0.1 $\theta_{12}$

The parameters for $\theta_{12}$ are passed in columnwise, with no transformations.

We apply a matrix normal prior, passing in a mean matrix, and the lower Cholesky decompositions for the two covariance matrices. The matrix normal is a standard parameterization.

## 0.2 $c$ and $u$

The average effects $\bar{c}$ and $\bar{u}$ are unconstrained, with normal priors. The priors on $\operatorname{sd} c$ and $\operatorname{sd} u$ are half-T. Hyperparameters are $\sigma_c$ and $\nu_c$ (similar for $u$). The Jacobians of the transformations are $\operatorname{sd} c$ and $\operatorname{sd} u$. Therefore,

$$\pi(\log \operatorname{sd}(c)) = \frac{2\Gamma(\frac{\nu_c+1}{2})}{\Gamma(\frac{\nu_c}{2})\sigma_c\sqrt{\nu_c \pi}} \left[1 + \frac{1}{\nu_c}\left(\frac{\operatorname{sd}(c)}{\sigma_c}\right)^2\right]^{-\frac{\nu_c+1}{2}} \operatorname{sd}(c) \tag{1}$$

$$\pi(\log \operatorname{sd}(u)) = \frac{2\Gamma(\frac{\nu_u+1}{2})}{\Gamma(\frac{\nu_u}{2})\sigma_u\sqrt{\nu_u \pi}} \left[1 + \frac{1}{\nu_u}\left(\frac{\operatorname{sd}(u)}{\sigma_u}\right)^2\right]^{-\frac{\nu_u+1}{2}} \operatorname{sd}(u) \tag{2}$$

The prior for each $c_j$ is normal, with mean $\bar{c}$ and standard deviation $\operatorname{sd}(c)$. To operationalize this, we immediately transform using the offsets

$$c_j = \operatorname{sd}_c\,(c_{\text{off}} + \bar{c}) \tag{3}$$

$$u_j = \operatorname{sd}_u\,(u_{\text{off}} + \bar{u}) \tag{4}$$

$$\tag{5}$$

This transformation lets us give each $c_{\text{off}}$ and $u_{\text{off}}$ a standard normal prior (mean=0, sd=1).

## 0.3 $\phi$

The coefficient matrix $\phi$ is passed in columnwise, with no transformation. It is included only if we are including the $H$ matrix in the model.

We apply a matrix normal prior, passing in a mean matrix, and the lower Cholesky decompositions for the two covariance matrices. The matrix normal is a standard parameterization.

## 0.4 $\delta$

We pass in $\operatorname{logit}\delta$, and transform so

$$\delta = \frac{\exp(\operatorname{logit}\delta)}{1 + \exp(\operatorname{logit}\delta)} \tag{6}$$

The prior on $\delta$ is a beta distribution with parameters $a_\delta$ and $b_\delta$. The Jacobian of the transformtion is $d\delta = \delta(1-\delta)$. Therefore,

$$\pi(\operatorname{logit}\delta) = \frac{\Gamma(a_\delta + b_\delta)}{\Gamma(a_\delta)\Gamma(b_\delta)}\delta^{a_\delta}(1-\delta)^{b_\delta} \tag{7}$$

Note that when taking the log of this prior, $\log\delta = \operatorname{logit}\delta - \log(1+\exp(\operatorname{logit}\delta))$, and $\log(1-\delta) = -\log(1+\exp(\operatorname{logit}\delta))$. This can be useful for avoiding numerical issues that would result from computing unnecessary logarithms.

### 0.5 $V_1$ and $V_2$

The covariance matrices $V_1$ and $V_2$ are structured similarly, so we will consider a general matrix $V$. We will let $V$ take a factor-analytic structure, where $x$ is a matrix of factors and $\Sigma$ is a diagonal matrix with all positive elements. We construct the matrix as

$$V = xx' + \Sigma \tag{8}$$

Let's start with $\Sigma$, and let the $i^{th}$ element of the diagonal be $\Sigma_{ii}$. The prior on each $\Sigma_{ii}$ is half-T. Since we are passing in $\log\Sigma_{ii}$ instead, we need to multiply each half-T density by $\Sigma_{ii}$ (the Jacobian of the transformation).

We arrange $x$ to have each column be a factor. For identification (need a reference for this), the upper triangle of $x$ is zero, and the elements of the diagonal are all positive. Therefore, if $x$ has $k$ rows and $n$ columns, there are only $kn - \frac{1}{2}n(n+1)$ unconstrained parameters and $\frac{1}{2}n(n-1)$ positive parameters. The unconstrained parameters all have T priors, and the positive ones have half-T priors. Only the half-T densities need to be multiplied by the Jacobian of the transformation.

When passing in the parameters, $\log\operatorname{diag}\Sigma$ comes first, followed by the elements of $x$, column-wise.

If there are no factors, then $V = \Sigma$.

### 0.6 W

Now, it gets fun. We partition $W$ so the upper left corner is $\alpha W_1$ and the lower right corner is $W_2$. The upper right and lower left corners are all zero. $W_2$ has the same factor-analytic structure as $V_1$ and $V_2$.

For identification, all elements of the diagonal of $W_1$ must be the same. Therefore, we treat $\alpha W_1$ as a scaled correlation matrix, where $\alpha > 0$ is the scale parameter is $W_1$ is a symmetric, positive definite matrix will all ones on the diagonal. The prior on $\log\alpha$ is a half-T prior, multiplied by $\alpha$ (the Jacobian).

We place an LKJ prior on $W_1$, with parameter $\eta$ (see LKJ paper and Stan manual). $W_1$ has $J+1$ rows/columns.

$$\pi(W_1) = C|W_1|^{\eta-1} \tag{9}$$

where

$$C = 2^{\sum_{i=1}^{J}(2\eta-2+J+1-i)(J+1-i)} \prod_{i=1}^{J} \left[ \mathbb{B}\left( \eta + \frac{J-i}{2}, \eta + \frac{J-i}{2} \right) \right]^{J+1-i} \tag{10}$$

Let $y$ be the vector of the $d = \binom{J+1}{2}$ unconstrained parameters, arranged columnwise in a lower triangular matrix. Let $z_{ij} = \tanh(y_{ij})$ (this is a Fisher transformation). Following the procedure in the Stan manual, construct another lower triangular matrix $x$ as follows:

1. $x_{11} = 1$
2. For $i = 2 \ldots J+1$, $x_{i1} = z_{i1}$ (copy first column)
3. For $i = 2 \ldots J+1$, $x_{ii} = \prod_{k=1}^{i-1} \sqrt{1 - z_{ik}^2}$ (diagonal elements)
4. For $j = 2 \ldots J+1$ and $i = j+1 \ldots J+1$, $x_{ij} = z_{ij} \prod_{k=1}^{j-1} \sqrt{1 - z_{ik}^2}$ (remaining off-diagonal elements)

This transformation will ensure that $W_1 = xx'$ is a valid correlation matrix.

Through the magic of Mathematica, we get the following Jacobian of the transformation

$$\mathcal{J}_{W_1} = \prod_{j=1}^{d-1} \prod_{i=j+1}^{d} \left[ \operatorname{sech}(y_{ij}) \right]^{d-j+1} \tag{11}$$

$$= \prod_{j=1}^{d-1} \prod_{i=j+1}^{d} \left[ 1 - z_{ij}^2 \right]^{\frac{d-j+1}{2}} \tag{12}$$

The operator sech is the hyperbolic secant. The second line comes from the identity $\operatorname{sech}^2(x) = 1 - \tanh^2(x)$.

Mathematica also tells us that

$$|W_1| = \prod_{j=1}^{d-1} \prod_{i=j+1}^{d} \left[ 1 - z_{ij}^2 \right] \tag{13}$$

If $\eta = 1$, the distribution is uniform over all correlation matrices. For $\eta > 1$, there is a mode at the identity matrix, and for $\eta < 1$ there is a trough.