BIOMETRICS 000, 000–000 DOI: 000
000 0000

Web-based Supplementary Materials for A Space-time Skew-t Model for Threshold Exceedances by Morris, Reich, Thibaud, and Cooley

Samuel A Morris[1,*], Brian J Reich[1], Emeric Thibaud[2], and Daniel Cooley[2]

[1]Department of Statistics, North Carolina State University, Raleigh, North Carolina, U.S.A.

[2]Department of Statistics, Colorado State University, Fort Collins, Colorado, U.S.A.

*_email:_ samorris@ncsu.edu

This paper has been submitted for consideration for publication in *Biometrics*

Web Appendix A. MCMC details

The MCMC sampling for the model 4 is done using R (http://www.r-project.org). Whenever possible, we select conjugate priors (see Appendix Web Appendix B); however, for some of the parameters, no conjugate prior distributions exist. For these parameters, we use a random walk Metropolis-Hastings update step. In each Metropolis-Hastings update, we tune the algorithm during the burn-in period to give acceptance rates near 0.40.

Spatial knot locations

For each day, we update the spatial knot locations, $\mathbf{w}_1, \ldots, \mathbf{w}_K$, using a Metropolis-Hastings block update. Because the spatial domain is bounded, we generate candidate knots using the transformed knots $\mathbf{w}_1^*, \ldots, \mathbf{w}_K^*$ (see section 3.3) and a random walk bivariate Gaussian candidate distribution

$$\mathbf{w}_k^{*(c)} \sim \text{N}(\mathbf{w}_k^{*(r-1)}, s^2 I_2)$$

where $\mathbf{w}_k^{*(r-1)}$ is the location for the transformed knot at MCMC iteration $r-1$, s is a tuning parameter, and I_2 is an identity matrix. After candidates have been generated for all K knots, the acceptance ratio is

$$R = \left\{ \frac{l[Y_t(\mathbf{s}|\mathbf{w}_1^{(c)}, \ldots, \mathbf{w}_K^{(c)}, \ldots)]}{l[Y_t(\mathbf{s}|\mathbf{w}_1^{(r-1)}, \ldots, \mathbf{w}_K^{(r-1)}, \ldots)]} \right\} \times \left\{ \frac{\prod_{k=1}^K \phi(\mathbf{w}_k^{(c)})}{\prod_{k=1}^K \phi(\mathbf{w}_k^{(r-1)})} \right\} \times \left\{ \frac{\prod_{k=1}^K p(\mathbf{w}_k^{*(c)})}{\prod_{k=1}^K p(\mathbf{w}_k^{*(r-1)})} \right\}$$

where l is the likelihood given in (18), and $p(\cdot)$ is the prior either taken from the time series given in (3.3) or assumed to be uniform over $\mathcal{D}$. The candidate knots are accepted with probability $\min\{R, 1\}$.

Spatial random effects

If there is no temporal dependence amongst the observations, we use a Gibbs update for z_{tk}, and the posterior distribution is given in Web Appendix B. If there is temporal dependence amongst the observations, then we update z_{tk} using a Metropolis-Hastings update. Because this model uses $|z_{tk}|$, we generate candidate random effects using the z_{tk}^* (see Section 3.3) and a random walk

Gaussian candidate distribution

$$z_{tk}^{*\,(c)} \sim \mathrm{N}(z_{tk}^{*\,(r-1)}, s^2)$$

where $z_{tk}^{*\,(r-1)}$ is the value at MCMC iteration $r-1$, and s is a tuning parameter. The acceptance ratio is

$$R = \left\{ \frac{l[Y_t(\mathbf{s})|z_{tk}^{(c)}, \ldots]}{l[Y_t(\mathbf{s})|z_{tk}^{(r-1)}]} \right\} \times \left\{ \frac{p[z_{tk}^{(c)}]}{p[z_{tk}^{(r-1)}]} \right\}$$

where $p[\cdot]$ is the prior taken from the time series given in Section 3.3. The candidate is accepted with probability $\min\{R, 1\}$.

Variance terms

When there is more than one site in a partition, then we update σ^2_{tk} using a Metropolis-Hastings update. First, we generate a candidate for σ^2_{tk} using an IG$(a^*/s, b^*/s)$ candidate distribution in an independence Metropolis-Hastings update where $a^* = (n_{tk}+1)/2+a$, $b^* = [Y_{tk}^T \Sigma_{tk}^{-1} Y_{tk} + z_{tk}^2]/2+b$, n_{tk} is the number of sites in partition k on day t, and Y_{tk} and Σ_{tk}^{-1} are the observations and precision matrix for partition k on day t. The acceptance ratio is

$$R = \left\{ \frac{l[Y_t(\mathbf{s})|\sigma_{tk}^{2\,(c)}, \ldots]}{l[Y_t(\mathbf{s})|\sigma_{tk}^{2\,(r-1)}]} \right\} \times \left\{ \frac{l[z_{tk}|\sigma_{tk}^{2\,(c)}, \ldots]}{l[z_{tk}|\sigma_{tk}^{2\,(r-1)}, \ldots]} \right\} \times \left\{ \frac{p[\sigma_{tk}^{2\,(c)}]}{p[\sigma_{tk}^{2\,(r-1)}]} \right\} \times \left\{ \frac{c[\sigma_{tk}^{2\,(r-1)}]}{c[\sigma_{tk}^{2\,(c)}]} \right\}$$

where $p[\cdot]$ is the prior either taken from the time series given in Section 3.3 or assumed to be IG(a, b), and $c[\cdot]$ is the candidate distribution. The candidate is accepted with probability $\min\{R, 1\}$.

Spatial covariance parameters

We update the three spatial covariance parameters, $\log(\rho)$, $\log(\nu)$, γ, using a Metropolis-Hastings block update step. First, we generate a candidate using a random walk Gaussian candidate distribution

$$\log(\rho)^{(c)} \sim \mathrm{N}(\log(\rho)^{(r-1)}, s^2)$$

where $\log(\rho)^{(r-1)}$ is the value at MCMC iteration $r-1$, and s is a tuning parameter. Candidates are generated for $\log(\nu)$ and γ in a similar fashion. The acceptance ratio is

$$R = \left\{ \frac{\prod_{t=1}^{T} l[Y_t(\mathbf{s}) | \rho^{(c)}, \nu^{(c)}, \gamma^{(c)}, \ldots]}{\prod_{t=1}^{T} l[Y_t(\mathbf{s}) | \rho^{(r-1)}, \nu^{(r-1)}, \gamma^{(r-1)}, \ldots]} \right\} \times \left\{ \frac{p[\rho^{(c)}]}{p[\rho^{(r-1)}]} \right\} \times \left\{ \frac{p[\nu^{(c)}]}{p[\nu^{(r-1)}]} \right\} \times \left\{ \frac{p[\gamma^{(c)}]}{p[\gamma^{(r-1)}]} \right\}.$$

All three candidates are accepted with probability $\min\{R, 1\}$.

Web Appendix B. Posterior distributions

Conditional posterior of $z_{tk} \mid \ldots$

If knots are independent over days, then the conditional posterior distribution of $|z_{tk}|$ is conjugate. For simplicity, drop the subscript t, let $\tilde{z}_{tk} = |z_{tk}|$, and define

$$R(\mathbf{s}) = \begin{cases} Y(\mathbf{s}) - X(\mathbf{s})\beta & s \in P_l \\ \\ Y(\mathbf{s}) - X(\mathbf{s})\beta - \lambda \tilde{z}(\mathbf{s}) & s \notin P_l \end{cases}$$

Let

$$\begin{aligned} R_1 &= \text{the vector of } R(\mathbf{s}) \text{ for } s \in P_l \\ R_2 &= \text{the vector of } R(\mathbf{s}) \text{ for } s \notin P_l \\ \Omega &= \Sigma^{-1}. \end{aligned}$$

Then

$$\begin{aligned} \pi(z_l | \ldots) &\propto \exp\left\{ -\frac{1}{2} \left[\begin{pmatrix} R_1 - \lambda \tilde{z}_l \mathbf{1} \\ R_2 \end{pmatrix}^T \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix} \begin{pmatrix} R_1 - \lambda \tilde{z}_l \mathbf{1} \\ R_2 \end{pmatrix} + \frac{\tilde{z}_l^2}{\sigma_l^2} \right] \right\} I(z_l > 0) \\ &\propto \exp\left\{ -\frac{1}{2} \left[\Lambda_l \tilde{z}_l^2 - 2\mu_l \tilde{z}_l \right] \right\} \end{aligned}$$

where

$$\begin{aligned} \mu_l &= \lambda (R_1^T \Omega_{11} + R_2^T \Omega_{21}) \mathbf{1} \\ \Lambda_l &= \lambda^2 \mathbf{1}^T \Omega_{11} \mathbf{1} + \frac{1}{\sigma_l^2}. \end{aligned}$$

Then $\tilde{Z}_l | \ldots \sim N_{(0,\infty)}(\Lambda_l^{-1}\mu_l, \Lambda_l^{-1})$

Conditional posterior of $\beta \mid \ldots$

Let $\beta \sim \mathrm{N}_p(0, \Lambda_0)$ where Λ_0 is a precision matrix. Then

$$\begin{aligned}
\pi(\beta \mid \ldots) &\propto \exp\left\{-\frac{1}{2}\beta^T\Lambda_0\beta - \frac{1}{2}\sum_{t=1}^{T}[\mathbf{Y}_t - X_t\beta - \lambda|z_t|]^T\Omega[\mathbf{Y}_t - X_t\beta - \lambda|z_t|]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\beta^T\Lambda_\beta\beta - 2\sum_{t=1}^{T}[\beta^T X_t^T\Omega(\mathbf{Y}_t - \lambda|z_t|)]\right]\right\} \\
&\propto \mathrm{N}(\Lambda_\beta^{-1}\mu_\beta, \Lambda_\beta^{-1})
\end{aligned}$$

where

$$\begin{aligned}
\mu_\beta &= \sum_{t=1}^{T}\left[X_t^T\Omega(\mathbf{Y}_t - \lambda|z_t|)\right] \\
\Lambda_\beta &= \Lambda_0 + \sum_{t=1}^{T} X_t^T\Omega X_t.
\end{aligned}$$

Conditional posterior of $\sigma^2 \mid \ldots$

In the case where $L = 1$ and temporal dependence is negligible, then σ^2 has a conjugate posterior distribution. Let $\sigma_t^2 \overset{iid}{\sim} \mathrm{IG}(\alpha_0, \beta_0)$. For simplicity, drop the subscript t. Then

$$\begin{aligned}
\pi(\sigma^2 \mid \ldots) &\propto (\sigma^2)^{-\alpha_0 - 1/2 - n/2 - 1}\exp\left\{-\frac{\beta_0}{\sigma^2} - \frac{|z|^2}{2\sigma^2} - \frac{(\mathbf{Y} - \boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{Y} - \boldsymbol{\mu})}{2\sigma^2}\right\} \\
&\propto (\sigma^2)^{-\alpha_0 - 1/2 - n/2 - 1}\exp\left\{-\frac{1}{\sigma^2}\left[\beta_0 + \frac{|z|^2}{2} + \frac{1}{2}(\mathbf{Y} - \boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{Y} - \mu)\right]\right\} \\
&\propto \mathrm{IG}(\alpha^*, \beta^*)
\end{aligned}$$

where

$$\begin{aligned}
\alpha^* &= \alpha_0 + \frac{1}{2} + \frac{n}{2} \\
\beta^* &= \beta_0 + \frac{|z|^2}{2} + \frac{1}{2}(\mathbf{Y} - \boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{Y} - \boldsymbol{\mu}).
\end{aligned}$$

In the case that $L > 1$, a random walk Metropolis Hastings step will be used to update σ_{lt}^2.

Conditional posterior of $\lambda \mid \ldots$

For convergence purposes we model $\lambda = \lambda_1 \lambda_2$ where

$$\lambda_1 = \begin{cases} +1 & \text{w.p.}0.5 \\ -1 & \text{w.p.}0.5 \end{cases} \tag{1}$$

$$\lambda_2^2 \sim IG(\alpha_\lambda, \beta_\lambda). \tag{2}$$

$$\tag{3}$$

Then

$$\begin{aligned} \pi(\lambda_2 \mid \ldots) &\propto \lambda_2^{2(-\alpha_\lambda - 1)} \exp\left\{-\frac{\beta_\lambda}{\lambda_2^2}\right\} \prod_{t=1}^{T} \prod_{k=1}^{K} \frac{1}{\lambda_2} \exp\left\{-\frac{z_{tk}^2}{2\lambda_2^2 \sigma_{tk})^2}\right\} \\ &\propto \lambda_2^{2(-\alpha_\lambda - kt - 1)} \exp\left\{-\frac{1}{\lambda_2^2}\left[\beta_\lambda + \frac{z^2}{2\sigma_{tk}^2}\right]\right\} \end{aligned}$$

Then $\lambda_2 \mid \ldots \sim IG\left(\alpha_\lambda + kt, \beta_\lambda + \frac{z^2}{2\sigma_{tk}^2}\right)$

Web Appendix C. Proof that $\lim_{h \to \infty} \pi(h) = 0$

Let c be the midpoint of $\mathbf{s}_1$ and $\mathbf{s}_2$. Define A as the circle centered at c with radius $h/2$ where $h = ||\mathbf{s}_1 - \mathbf{s}_2||$ is the distance between sites $\mathbf{s}_1$ and $\mathbf{s}_2$. Consider a homogeneous spatial Poisson process over A with intensity given by

$$\mu(A) = \lambda_{PP}|A| = \lambda_{PP}\pi\left(\frac{h}{2}\right)^2 = \lambda_{PPA}^* h^2.$$

Consider a partition of A into four regions, B_1, B_2, R_1, R_2 as seen in Web Figure 1.

[Figure 1 about here.]

Let N_j be the number of knots in $B_j, j = 1, 2$. Then

$$P(\mathbf{s}_1 \in P_i, \mathbf{s}_2 \in P_{j \neq i}) \geqslant P(N_1 > 0, N_2 > 0) \tag{4}$$

since knots in both B_1 and B_2 is sufficient, but not necessary, to ensure that $\mathbf{s}_1$ and $\mathbf{s}_2$ are in different partition sets. By definition of a Poission process, N_1 and N_2 are independent and thus

$P(N_1 > 0, N_2 > 0) = P(N_1 > 0)^2$, and the intensity measure over B_1 is given by

$$\mu(B_1) = \lambda_{PP}|B_1| = \lambda_{PP}\frac{h^2}{4}\left(\frac{2\pi}{3} - \frac{\sqrt{3}}{2}\right)$$

$$= \lambda^*_{PPB1}h^2. \tag{5}$$

So,

$$P(\mathbf{s}_1 \in P_i, \mathbf{s}_2 \in P_{j\neq i}) >= P(N_1 > 0)^2 = [1 - P(N_1 = 0)]^2 = [1 - \exp\left(-\lambda^*_{PPB1}h^2\right)]^2 \tag{6}$$

which goes to 1 as h goes to infinity.

Web Appendix D. Skew-t distribution

Univariate skew-t distribution

We say that Y follows a univariate extended skew-t distribution with location $\xi \in \mathcal{R}$, scale $\omega > 0$, skew parameter $\alpha \in \mathcal{R}$, and degrees of freedom ν if has distribution function

$$f_{\text{EST}}(y) = 2f_T(z;\nu)F_T\left[\alpha z\sqrt{\frac{\nu+1}{\nu+z^2}};\nu+1\right] \tag{7}$$

where $f_T(t;\nu)$ is a univariate Student's t with ν degrees of freedom, $F_T(t;\nu) = P(T < t)$, and $z = (y-\xi)/\omega$.

Multivariate skew-t distribution

If $\mathbf{Z} \sim \text{ST}_d(0, \bar{\mathbf{\Omega}}, \boldsymbol{\alpha}, \eta)$ is a d-dimensional skew-t distribution, and $\mathbf{Y} = \xi + \boldsymbol{\omega}\mathbf{Z}$, where $\boldsymbol{\omega} = \text{diag}(\omega_1, \ldots, \omega_d)$, then the density of Y at y is

$$f_y(\mathbf{y}) = det(\boldsymbol{\omega})^{-1}f_z(\mathbf{z}) \tag{8}$$

where

$$f_z(\mathbf{z}) = 2t_d(\mathbf{z};\bar{\mathbf{\Omega}},\eta)T\left[\boldsymbol{\alpha}^T\mathbf{z}\sqrt{\frac{\eta+d}{\nu+Q(\mathbf{z})}};\eta+d\right] \tag{9}$$

$$\mathbf{z} = \boldsymbol{\omega}^{-1}(\mathbf{y}-\xi) \tag{10}$$

where $t_d(\mathbf{z}; \bar{\boldsymbol{\Omega}}, \eta)$ is a d-dimensional Student's t-distribution with scale matrix $\bar{\boldsymbol{\Omega}}$ and degrees of freedom η, $Q(z) = \mathbf{z}^T\bar{\Omega}^{-1}\mathbf{z}$ and $T(\cdot;\eta)$ denotes the univariate Student's t distribution function with η degrees of freedom (Azzalini and Capitanio, 2014).

Extremal dependence

For a bivariate skew-t random variable $\mathbf{Y} = [Y(\mathbf{s}), Y(\mathbf{t})]^T$, the $\chi(h)$ statistic (Padoan, 2011) is given by

$$\chi(h) = \bar{F}_{\text{EST}}\left\{\frac{[x_1^{1/\eta} - \varrho(h)]\sqrt{\eta+1}}{\sqrt{1-\varrho(h)^2}}; 0, 1, \alpha_1, \tau_1, \eta+1\right\} + \bar{F}_{\text{EST}}\left\{\frac{[x_2^{1/\eta} - \varrho(h)]\sqrt{\eta+1}}{\sqrt{1-\varrho(h)^2}}; 0, 1, \alpha_2, \tau_2, \eta+1\right\}, \tag{11}$$

where $\bar{F}_{\text{EST}}$ is the univariate survival extended skew-t function with zero location and unit scale, $\varrho(h) = \text{cor}[y(\mathbf{s}), y(\mathbf{t})]$, $\alpha_j = \alpha_i\sqrt{1-\varrho^2}$, $\tau_j = \sqrt{\eta+1}(\alpha_j+\alpha_i\varrho)$, and $x_j = F_T(\bar{\alpha}_i\sqrt{\eta+1}; 0, 1, \eta)/F_T(\bar{\alpha}_j\sqrt{\eta+1};$ with $j = 1, 2$ and $i = 2, 1$ and where $\bar{\alpha}_j = (\alpha_j + \alpha_i\varrho)/\sqrt{1+\alpha_i^2[1-\varrho(h)^2]}$.

Proof that $\lim_{h\to\infty}\chi(h) > 0$

Consider the bivariate distribution of $\mathbf{Y} = [Y(\mathbf{s}), Y(\mathbf{t})]^T$, with $\varrho(h)$ given by (3). So, $\lim_{h\to\infty}\varrho(h) = 0$. Then

$$\lim_{h\to\infty}\chi(h) = \bar{F}_{\text{EST}}\left\{\sqrt{\eta+1}; 0, 1, \alpha_1, \tau_1, \eta+1\right\} + \bar{F}_{\text{EST}}\left\{\sqrt{\eta+1}; 0, 1, \alpha_2, \tau_2, \eta+1\right\}. \tag{12}$$

Because the extended skew-t distribution is not bounded above, for all $\bar{F}_{\text{EST}}(x) = 1 - F_{\text{EST}(x)} > 0$ for all $x < \infty$. Therefore, for a skew-t distribution, $\lim_{h\to\infty}\chi(h) > 0$.

Web Appendix E. Simulation study pairwise difference results

The following tables show the methods that have significantly different Brier scores when using a Wilcoxon-Nemenyi-McDonald-Thompson test. In each column, different letters signify that the methods have significantly different Brier scores. For example, there is significant evidence to suggest that method 1 and method 4 have different Brier scores at $q(0.90)$, whereas there is not significant evidence to suggest that method 1 and method 2 have different Brier scores at $q(0.90)$. In each table group A represents the group with the lowest Brier scores. Groups are significant with a familywise error rate of $\alpha = 0.05$.

[Table 1 about here.]

[Table 2 about here.]

[Table 3 about here.]

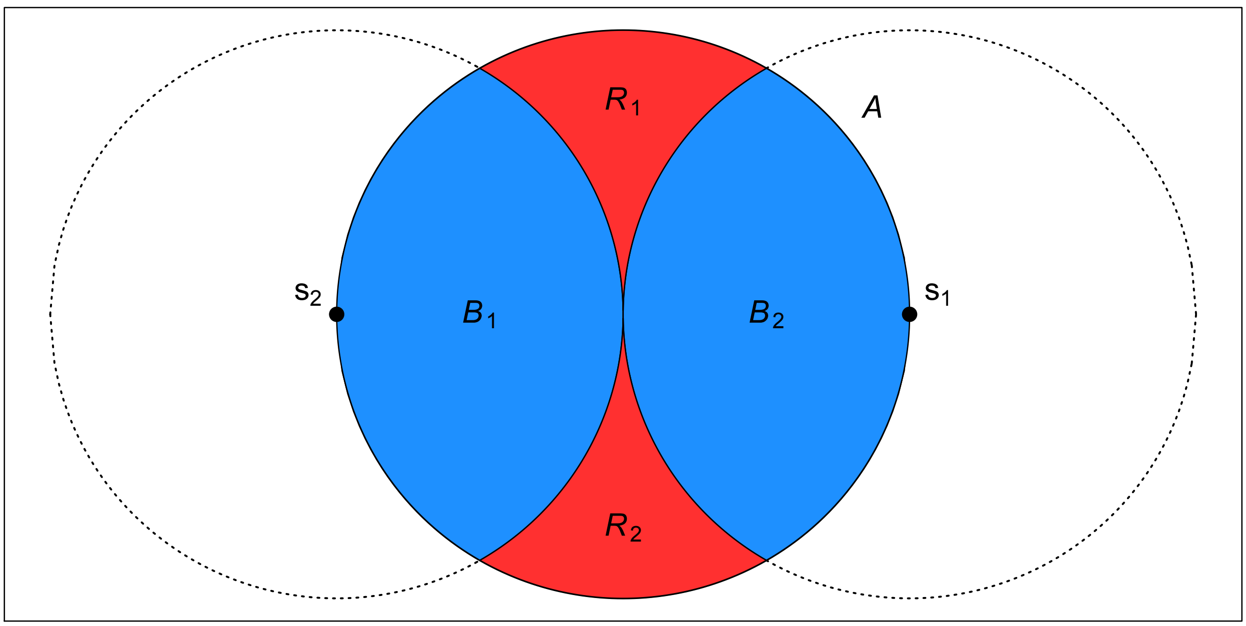
[Table 4 about here.]

[Table 5 about here.]

References

Azzalini, A. and Capitanio, A. (2014). *The Skew-Normal and Related Families*. Institute of Mathematical Statistics Monographs. Cambridge University Press.

Padoan, S. A. (2011). Multivariate extreme models based on underlying skew- and skew-normal distributions. *Journal of Multivariate Analysis* **102,** 977–991.



Web Figure 1. Illustration of the partition of A.

Web Table 1

Setting 1 – Gaussian marginal, $K = 1$ knot

	$q(0.90)$	$q(0.95)$	$q(0.98)$	$q(0.99)$
Method 1	A	A	A	A B
Method 2	A	A	A	A
Method 3	B	B	C	B
Method 4	A	A	A B	A B
Method 5	B	B	B C	A B
Method 6	C	C	D	C

Web Table 2

Setting 2 – Skew-t marginal, $K = 1$ knot

	$q(0.90)$	$q(0.95)$	$q(0.98)$	$q(0.99)$
Method 1	C	B	B C	B
Method 2	A	A	A	A
Method 3	B C	A B	A B	A B
Method 4	A B	B	B	A
Method 5	D	C	C	B
Method 6	E	D	D	C

Web Table 3

Setting 3 – Skew-t marginal, $K = 5$ knots

	$q(0.90)$	$q(0.95)$	$q(0.98)$	$q(0.99)$
Method 1	B	C	B	B
Method 2	B	C	B	B
Method 3	A	B	B	B
Method 4	A	A	A	A
Method 5	A	A	A	A
Method 6	C	D	C	C

Web Table 4
Setting 4 – Max-stable

	$q(0.90)$	$q(0.95)$	$q(0.98)$	$q(0.99)$
Method 1	A B	B	B	C
Method 2	B	B C	B	B C
Method 3	C D	C	B	B
Method 4	D	D	C	C
Method 5	C	C	B	B C
Method 6	A	A	A	A

Web Table 5

Setting 5 – Transformation below $T = q(0.80)$

	$q(0.90)$	$q(0.95)$	$q(0.98)$	$q(0.99)$
Method 1	C	B	C	C
Method 2	B	B	B	A B
Method 3	A	A	A	A
Method 4	B C	B	B	B C
Method 5	B	B	B C	C
Method 6	D	C	D	D