Modelling Shanghai Soil Properties with Finite Mixtures of S_{U} Johnson Distributions – Supplementary Material

Nikola Počuča[a], Michael P.B. Gallaugher[b], Paul D. McNicholas[a]

[a] *Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada*
[b] *Department of Statistical Science, Baylor University, Waco, Texas, USA*

1. Complete-Data Profile-Loglikelihood

There are a large number of parameters required to estimate in our model. It is favourable to reduce the dimensionality of parameter optimizations through the derivation of a profile log-likelihood. The work of [1] shows that profile log-likelihoods behave like ordinary likelihoods, in that they have a quadratic expansion, and therefore, can be used to fit maximum likelihood estimates. Spitzer [2] provides a primer on Box-Cox parameter estimations using profile log-likelihoods. Due to the similarities between both S_{U} and power transformations, we consider a similar approach as follows.

Let $\boldsymbol{\Phi} := (\boldsymbol{\Theta} \setminus \boldsymbol{\vartheta})$ designate all parameters that are not associated with the scale/shift of hyperbolic space. The maximum likelihood estimates $\hat{\boldsymbol{\Phi}}$ of Equation (7) in the main manuscript are derived as follows

$$n_g = \sum_{i=1}^{n} z_{ig}, \quad \hat{\pi}_g = \frac{\sum_{i=1}^{n} z_{ig}}{n_g}, \quad \hat{\boldsymbol{\mu}}_g = \frac{\sum_{i=1}^{n} z_{ig}\boldsymbol{h}(\boldsymbol{y}_i;\boldsymbol{\vartheta}_g)}{n_g},$$

$$\hat{\boldsymbol{\Sigma}}_g = \frac{\sum_{i=1}^{n} z_{ig}\left(\boldsymbol{h}(\boldsymbol{y}_i;\boldsymbol{\vartheta}_g) - \hat{\boldsymbol{\mu}}_g\right)\left(\boldsymbol{h}(\boldsymbol{y}_i;\boldsymbol{\vartheta}_g) - \hat{\boldsymbol{\mu}}_g\right)^{\top}}{n_g}. \tag{1}$$

Substituting maximum likelihood estimates of (1) into Equation (7) of the main manuscript yields the complete-profile log-likelihood for $\boldsymbol{\vartheta}$ as

$$
\begin{aligned}
l_c(\boldsymbol{\vartheta}; \boldsymbol{y}_1, \ldots, \boldsymbol{y}_n, \boldsymbol{z}, \hat{\boldsymbol{\Phi}}) = & \sum_{i=1}^{n} \sum_{g=1}^{G} z_{ig} \log(\hat{\pi}_g) \\
& - \frac{1}{2} \sum_{i=1}^{n} \sum_{g=1}^{G} z_{ig} \left(\boldsymbol{h}(\boldsymbol{y}_i; \boldsymbol{\vartheta}_g) - \hat{\boldsymbol{\mu}}_g \right)^\top \hat{\boldsymbol{\Sigma}}_g^{-1} \left(\boldsymbol{h}(\boldsymbol{y}_i; \boldsymbol{\vartheta}_g) - \hat{\boldsymbol{\mu}}_g \right)) \\
& - \frac{np}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^{n} \sum_{g=1}^{G} z_{ig} \log(|\hat{\boldsymbol{\Sigma}}_g|) - \sum_{i=1}^{n} \sum_{g=1}^{G} \sum_{j=1}^{p} z_{ig} \log(\delta_{jg}) \\
& - \frac{1}{2} \sum_{i=1}^{n} \sum_{g=1}^{G} \sum_{j=1}^{p} z_{ig} \log \left[\left(\frac{y_{ij} - \omega_{jg}}{\delta_{jg}} \right)^2 + 1 \right].
\end{aligned}
$$

We can show that the second term is free of $\boldsymbol{\vartheta}$ as follows:

$$
\begin{aligned}
& - \frac{1}{2} \sum_{i=1}^{n} \sum_{g=1}^{G} z_{ig} \left(\boldsymbol{h}(\boldsymbol{y}_i; \boldsymbol{\vartheta}_g) - \hat{\boldsymbol{\mu}}_g \right)^\top \hat{\boldsymbol{\Sigma}}_g^{-1} \left(\boldsymbol{h}(\boldsymbol{y}_i; \boldsymbol{\vartheta}_g) - \hat{\boldsymbol{\mu}}_g \right)) \\
= & - \frac{1}{2} \sum_{i=1}^{n} \sum_{g=1}^{G} z_{ig} \left(\boldsymbol{h}(\boldsymbol{y}_i; \boldsymbol{\vartheta}_g) - \hat{\boldsymbol{\mu}}_g \right)^\top \left[\frac{\sum_{i=1}^{n} z_{ig} \left(\boldsymbol{h}(\boldsymbol{y}_i; \boldsymbol{\vartheta}_g) - \hat{\boldsymbol{\mu}}_g \right) \left(\boldsymbol{h}(\boldsymbol{y}_i; \boldsymbol{\vartheta}_g) - \hat{\boldsymbol{\mu}}_g \right)^\top}{n_g} \right]^{-1} \\
& \times \left(\boldsymbol{h}(\boldsymbol{y}_i; \boldsymbol{\vartheta}_g) - \hat{\boldsymbol{\mu}}_g \right)) .
\end{aligned}
$$

Let us simplify the expression further, let $\boldsymbol{b}_{ig} = (\boldsymbol{h}(\boldsymbol{y}_i; \boldsymbol{\vartheta}_g) - \hat{\boldsymbol{\mu}}_g)$ allowing

$$
\begin{aligned}
& - \frac{1}{2} \sum_{i=1}^{n} \sum_{g=1}^{G} z_{ig} \boldsymbol{b}_{ig}^\top \left(\frac{\sum_{i=1}^{n} z_{ig} \boldsymbol{b}_{ig} \boldsymbol{b}_{ig}^\top}{n_g} \right)^{-1} \boldsymbol{b}_{ig} = - \frac{1}{2} \sum_{g=1}^{G} n_g \sum_{i=1}^{n} z_{ig} \mathrm{trace} \left[\boldsymbol{b}_{ig} \boldsymbol{b}_{ig}^\top \left(\sum_{i=1}^{n} z_{ig} \boldsymbol{b}_{ig} \boldsymbol{b}_{ig}^\top \right)^{-1} \right] \\
& = - \frac{1}{2} \sum_{g=1}^{G} n_g \mathrm{trace} \left[\sum_{i=1}^{n} z_{ig} \boldsymbol{b}_{ig} \boldsymbol{b}_{ig}^\top \left(\sum_{i=1}^{n} z_{ig} \boldsymbol{b}_{ig} \boldsymbol{b}_{ig}^\top \right)^{-1} \right] = - \frac{1}{2} \sum_{g=1}^{G} n_g \mathrm{trace} \left(\mathbf{I} \right) = - \frac{np}{2}.
\end{aligned}
$$

With this simplification, the complete-profile likelihood can be written as:

$$
\begin{aligned}
l_{\mathrm{cp}}(\boldsymbol{\vartheta};\boldsymbol{y}_1,\ldots,\boldsymbol{y}_n) = &\sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\log(\hat{\pi}_g) - \frac{np}{2}\log(2\pi) + \frac{np}{2}\sum_{g=1}^{G}\log(n_g) - \frac{np}{2} \\
&- \frac{1}{2}\sum_{i=1}^{n}\sum_{g=1}^{G} z_{ig}\text{logdet}\left(\sum_{i=1}^{n} z_{ig}\left(\boldsymbol{h}(\boldsymbol{y}_i;\boldsymbol{\vartheta}_g) - \hat{\boldsymbol{\mu}}_g\right)\left(\boldsymbol{h}(\boldsymbol{y}_i;\boldsymbol{\vartheta}_g) - \hat{\boldsymbol{\mu}}_g\right)^{\top}\right) \\
&- \sum_{i=1}^{n}\sum_{g=1}^{G}\sum_{j=1}^{p} z_{ig}\log(\delta_{jg}) - \frac{1}{2}\sum_{i=1}^{n}\sum_{g=1}^{G}\sum_{j=1}^{p} z_{ig}\log\left[\left(\frac{y_{ij}-\omega_{jg}}{\delta_{jg}}\right)^{2}+1\right].
\end{aligned}
$$

Notice that the first row of terms are strictly constants with respect to $\boldsymbol{\vartheta}$ and are deemed unnecessary to compute during the optimization procedure. By concentrating out the terms for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, we have reduced the optimization procedure down to strictly $2pG$ parameters.

2. Derivation of a Conditional S_{U} Distribution

Let $\boldsymbol{Y}$ emanate from a S_{U} distribution parametrized by $\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\omega}$, and $\boldsymbol{\Lambda}$. For some realization $\boldsymbol{y} \in \boldsymbol{Y}$, consider the conditional distribution of $y_j|\boldsymbol{y}_d$ where $\boldsymbol{y}_d$ corresponds to entries of $\boldsymbol{y}$ excluding y_j. We now show that the distribution of $y_j|\boldsymbol{y}_d$ is a univariate S_{U} Johnson distribution parametrized by $\tilde{\mu}, \tilde{\sigma}, \omega_j, \delta_j$ as follows. The join density function of $\boldsymbol{y} = (y_j, \boldsymbol{y}_d)$ is written as

$$
f(\boldsymbol{y};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\omega},\boldsymbol{\Lambda}) = \frac{\exp\left\{-\frac{1}{2}\left(\boldsymbol{h}(\boldsymbol{y};\boldsymbol{\vartheta}) - \boldsymbol{\mu}\right)^{\top}\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{h}(\boldsymbol{y};\boldsymbol{\vartheta}) - \boldsymbol{\mu}\right)\right\}}{(2\pi)^{\frac{p}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}\prod_{k=1}^{p}\left(\delta_k\sqrt{\left(\frac{y_k-\omega_k}{\delta_k}\right)^{2}+1}\right)} \tag{2}
$$

for $\boldsymbol{y} \in \mathbb{R}^p$. Within our derivation, and without loss of generality, we place index j as the first entry for convenience. Working with the denominator of

(2), we factor into two separate forms [using 3, 2.8.13] as

$$
\begin{aligned}
&(2\pi)^{\frac{p}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}\prod_{k=1}^{p}\left(\delta_k\sqrt{\left(\frac{y_k-\omega_k}{\delta_k}\right)^2+1}\right)|\boldsymbol{\Sigma}|^{\frac{1}{2}} \\
&=(2\pi)^{\frac{1}{2}}(2\pi)^{\frac{(p-1)}{2}}\left(\delta_j\sqrt{\left(\frac{y_j-\omega_j}{\delta_j}\right)^2+1}\right)\prod_{k=2}^{p}\left(\delta_k\sqrt{\left(\frac{y_k-\omega_k}{\delta_k}\right)^2+1}\right) \\
&|\boldsymbol{\Sigma}_{d,d}|^{\frac{1}{2}}|\sigma_{jj}^2-\boldsymbol{\Sigma}_{j,d}\boldsymbol{\Sigma}_{d,d}^{-1}\boldsymbol{\Sigma}_{d,j}|^{\frac{1}{2}} \\
&=(2\pi)^{\frac{1}{2}}|\sigma_{jj}^2-\boldsymbol{\Sigma}_{j,d}\boldsymbol{\Sigma}_{d,d}^{-1}\boldsymbol{\Sigma}_{d,j}|^{\frac{1}{2}}\left(\delta_j\sqrt{\left(\frac{y_j-\omega_j}{\delta_j}\right)^2+1}\right)(2\pi)^{\frac{(p-1)}{2}} \\
&\times\prod_{k=2}^{p}\left(\delta_k\sqrt{\left(\frac{y_k-\omega_k}{\delta_k}\right)^2+1}\right)|\boldsymbol{\Sigma}_{d,d}|^{\frac{1}{2}}. \qquad (3)
\end{aligned}
$$

Working with the numerator of (2), we can decompose the quadratic form of the exponent into two parts. Let $\boldsymbol{a}=\boldsymbol{h}(\boldsymbol{y};\boldsymbol{\vartheta})$ for ease of notation. Since $\boldsymbol{h}$ is a component-wise function, we can treat each entry within as $\boldsymbol{a}=\left[a_j,\boldsymbol{a}_d\right]$. Now, using [3, 2.8.27], we decompose our quadratic form as

$$
\begin{aligned}
(\boldsymbol{a}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{a}-\boldsymbol{\mu}) &= \left[a_j-\mu_j,\boldsymbol{a}_d-\boldsymbol{\mu}_d\right]\begin{bmatrix}\sigma_{jj}^2, & \boldsymbol{\Sigma}_{jd}\\ \boldsymbol{\Sigma}_{dj}, & \boldsymbol{\Sigma}_{dd}\end{bmatrix}^{-1}\begin{bmatrix}a_j-\mu_j\\ \boldsymbol{a}_d-\boldsymbol{\mu}_d\end{bmatrix} \\
&=(a_j-\mu_j-\boldsymbol{\Sigma}_{jd}\boldsymbol{\Sigma}_{dd}^{-1}(\boldsymbol{a}_d-\boldsymbol{\mu}_d))^\top(\sigma_{jj}^2-\boldsymbol{\Sigma}_{jd}\boldsymbol{\Sigma}_{dd}^{-1}\boldsymbol{\Sigma}_{dj})^{-1} \\
&\quad\times(a_j-\mu_j-\boldsymbol{\Sigma}_{jd}\boldsymbol{\Sigma}_{dd}^{-1}(\boldsymbol{a}_d-\boldsymbol{\mu}_d))+(\boldsymbol{a}_d-\boldsymbol{\mu}_d)^\top\boldsymbol{\Sigma}_{dd}^{-1}(\boldsymbol{a}_d-\boldsymbol{\mu}_d). \qquad (4)
\end{aligned}
$$

Given the preceding results, we derive the density of the conditional distribution $y_j|\boldsymbol{y}_d$ as

$$
\begin{aligned}
f(y_j|\boldsymbol{y}_d;\tilde{\mu},\tilde{\sigma},\omega_j,\delta_j) &= \frac{f(\boldsymbol{y};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\omega},\boldsymbol{\Lambda})}{f(\boldsymbol{y}_d;\boldsymbol{\mu}_d,\boldsymbol{\Sigma}_d,\boldsymbol{\omega}_d,\boldsymbol{\Lambda}_d)} \\
&=\frac{(2\pi)^{\frac{p-1}{2}}|\boldsymbol{\Sigma}_{dd}|^{\frac{1}{2}}\prod_{l=2}^{p}\left(\delta_l\sqrt{\left(\frac{y_l-\omega_l}{\delta_l}\right)^2+1}\right)}{(2\pi)^{\frac{p}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}\prod_{k=1}^{p}\left(\delta_k\sqrt{\left(\frac{y_k-\omega_k}{\delta_k}\right)^2+1}\right)}\frac{\exp\left\{-\frac{1}{2}\left(\boldsymbol{a}-\boldsymbol{\mu}\right)^\top\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{a}-\boldsymbol{\mu}\right)\right\}}{\exp\left\{-\frac{1}{2}(\boldsymbol{a}_d-\boldsymbol{\mu}_d)^\top\boldsymbol{\Sigma}_{dd}^{-1}(\boldsymbol{a}_d-\boldsymbol{\mu}_d)\right\}}. \qquad (5)
\end{aligned}
$$

Working with the left hand side of (5), and using (3), we can reduce the expression to

$$\frac{(2\pi)^{\frac{p-1}{2}}|\mathbf{\Sigma}_{dd}|^{\frac{1}{2}}\prod_{l=2}^{p}\left(\delta_l\sqrt{\left(\frac{y_l-\omega_l}{\delta_l}\right)^2+1}\right)}{(2\pi)^{\frac{p}{2}}|\mathbf{\Sigma}|^{\frac{1}{2}}\prod_{k=1}^{p}\left(\delta_k\sqrt{\left(\frac{y_k-\omega_k}{\delta_k}\right)^2+1}\right)}=\frac{1}{(2\pi)^{\frac{1}{2}}|\sigma_{jj}^2-\mathbf{\Sigma}_{j,d}\mathbf{\Sigma}_{d,d}^{-1}\mathbf{\Sigma}_{d,j}|^{\frac{1}{2}}\left(\delta_j\sqrt{\left(\frac{y_j-\omega_j}{\delta_j}\right)^2+1}\right)}. \tag{6}$$

Working with the exponential terms at right hand side of (5) and, using (4), we have that

$$\begin{aligned}
&(\boldsymbol{a}-\boldsymbol{\mu})^\top\mathbf{\Sigma}^{-1}(\boldsymbol{a}-\boldsymbol{\mu})-(\boldsymbol{a}_d-\boldsymbol{\mu}_d)^\top\mathbf{\Sigma}_{dd}^{-1}(\boldsymbol{a}_d-\boldsymbol{\mu}_d)\\
&=(a_j-\mu_j-\mathbf{\Sigma}_{jd}\mathbf{\Sigma}_{dd}^{-1}(\boldsymbol{a}_d-\boldsymbol{\mu}_d))^\top(\sigma_{jj}^2-\mathbf{\Sigma}_{jd}\mathbf{\Sigma}_{dd}^{-1}\mathbf{\Sigma}_{dj})^{-1}(a_j-\mu_j-\mathbf{\Sigma}_{jd}\mathbf{\Sigma}_{dd}^{-1}(\boldsymbol{a}_d-\boldsymbol{\mu}_d))\\
&\quad+(\boldsymbol{a}_d-\boldsymbol{\mu}_d)^\top\mathbf{\Sigma}_{dd}^{-1}(\boldsymbol{a}_d-\boldsymbol{\mu}_d)-(\boldsymbol{a}_d-\boldsymbol{\mu}_d)^\top\mathbf{\Sigma}_{dd}^{-1}(\boldsymbol{a}_d-\boldsymbol{\mu}_d)\\
&=(a_j-\mu_j-\mathbf{\Sigma}_{jd}\mathbf{\Sigma}_{dd}^{-1}(\boldsymbol{a}_d-\boldsymbol{\mu}_d))^\top(\sigma_{jj}^2-\mathbf{\Sigma}_{jd}\mathbf{\Sigma}_{dd}^{-1}\mathbf{\Sigma}_{dj})^{-1}(a_j-\mu_j-\mathbf{\Sigma}_{jd}\mathbf{\Sigma}_{dd}^{-1}(\boldsymbol{a}_d-\boldsymbol{\mu}_d)).
\end{aligned} \tag{7}$$

Combining results (6) and (7), we have that $\tilde{\mu}=\mu_j+\mathbf{\Sigma}_{jd}\mathbf{\Sigma}_{dd}^{-1}(\boldsymbol{a}_d-\boldsymbol{\mu}_d)$, $\tilde{\sigma}^2=\sigma_{jj}^2-\mathbf{\Sigma}_{jd}\mathbf{\Sigma}_{dd}^{-1}\mathbf{\Sigma}_{dj}$, and

$$\begin{aligned}
f(y_j|\boldsymbol{y}_d;\tilde{\mu},\tilde{\sigma},\omega_j,\delta_j)&=\frac{\exp\left\{-\frac{1}{2}(a_j-\tilde{\mu})^\top(\tilde{\sigma}^2)^{-1}(a_j-\tilde{\mu})\right\}}{(2\pi)^{\frac{1}{2}}|\tilde{\sigma}^2|^{\frac{1}{2}}\left(\delta_j\sqrt{\left(\frac{y_j-\omega_j}{\delta_j}\right)^2+1}\right)}\\
&=\frac{\exp\left\{-\frac{1}{2}\left(\frac{h(y_j;\omega_j,\delta_j)-\tilde{\mu}}{\tilde{\sigma}}\right)^2\right\}}{(2\pi)^{\frac{1}{2}}\tilde{\sigma}\left(\delta_j\sqrt{\left(\frac{y_j-\omega_j}{\delta_j}\right)^2+1}\right)},
\end{aligned}$$

which is the density of a univariate S_{U} Johnson distribution with hyperbolic parameters ω_j and δ_j, and Gaussian parameters $\tilde{\mu}$ and $\tilde{\sigma}$. Notice that the conditioning on $\boldsymbol{y}_d$ only adjusts the Gaussian parameters giving rise to the same joint distribution as $\boldsymbol{y}$.

References

[1] S. A. Murphy, A. W. Van der Vaart, On profile likelihood, Journal of the American Statistical Association 95 (2000) 449–465.

[2] J. J. Spitzer, A primer on Box-Cox estimation, The Review of Economics and Statistics (1982) 307–313.

[3] S. B. Dennis, Matrix Mathematics: Theory, Facts and Formulas, Princeton University Press, Princeton, NJ 42 (2009) 1139.