Semiparametric Gaussian Variance-Mean Mixtures for Heavy-Tailed and Skewed Data

Kai Cui
Department of Statistical Science, Duke University

October 29, 2012

Appendix

A Proofs

Proof *of Lemma 2.2.* Given p-dimensional observation y from a Gaussian variance mean mixtures described in (1.1). Let f_G^* denote the pdf of the Gaussian variance mean mixture with mixing distribution G^*, and let f_G denote the pdf of the Gaussian variance mixture ($\gamma = 0$) with mixing distribution G . Without loss of generality, we considered the case with $\mu = 0$ (otherwise, a linear transformation $\mu + y$ will complete the proof).

$$\begin{aligned} f(y) &\propto \int_0^\infty V^{-p/2} \exp\left(-\frac{1}{2}y'\frac{1}{V}\Sigma^{-1}yg(V)dV\right) \\ &\propto \int_0^\infty V^{-p/2} \exp\left(-\frac{1}{2}y'\frac{1}{V}\Sigma^{-1}y\right) \exp\left(-\frac{1}{2}\gamma'\Sigma^{-1}\gamma V\right) \exp\left(\frac{1}{2}\gamma'\Sigma^{-1}\gamma V\right) g(V)dV \end{aligned}$$

Let $\delta = \frac{1}{2}\gamma'\Sigma^{-1}\gamma$, with $\delta = \frac{1}{2}\gamma^2 \geq 0$ in the univariate case, and define

$$g^*(V) = \frac{g(V)\exp(\delta V)}{\varphi(\delta)}$$

where g and g^* are densities for G and G^* respectively, and φ is the moment generating function for g, with $\varphi(\delta) < \infty$. Plug g^* in, and we can see that

$$f_{g^*}(y) = \frac{\exp(\gamma y)}{\varphi(\gamma^2/2)} f_g(y)$$

.

B Supplemental Tables and Figures

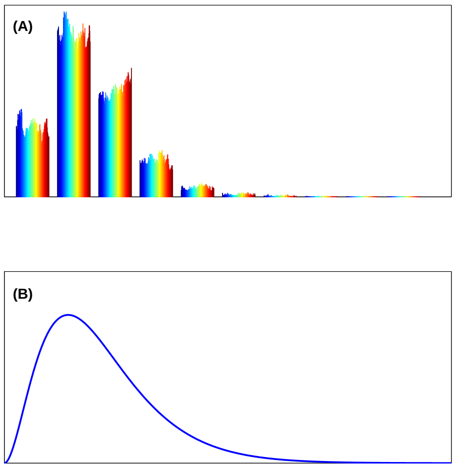

Figure S1: A comparison between the true mixing distribution (Gamma(3,1), panel (B)) and histograms of 100 reconstructed mixing distributions (panel (A)) show significant similarity.

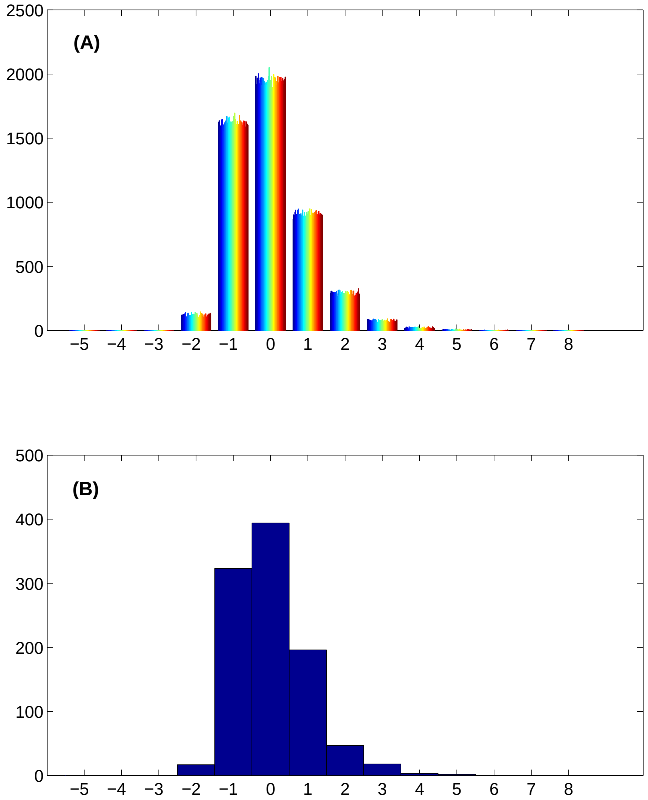

Figure S2: Comparison of Reconstructed Distribution of Data with the Observed Distribution. Figure (A) on the top shows histograms of 200 reconstructed dataset, each consisting of 5000 data points. And figure (B) on the bottom is the histogram of observed data.

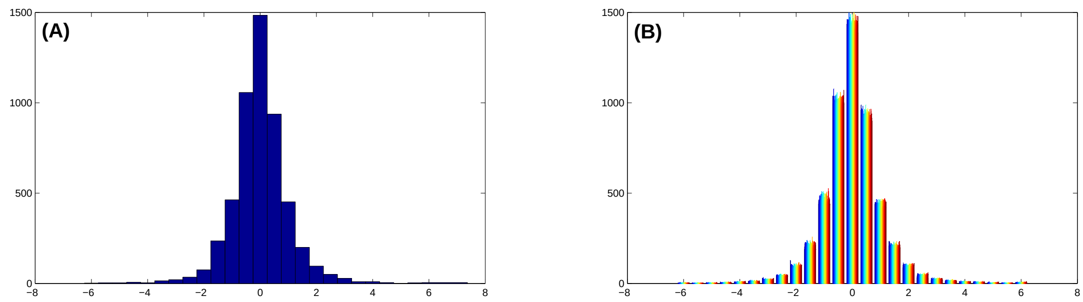

Figure S3: Fitting S&P 500 index return via univariate NVMM. Panel (A) shows the histogram of 200 reconstructed dataset (each consisting of 5470 observations) based on posterior samples. Panel (B) shows the real S&P 500 return from 01/02/1990 to 09/13/2011, totally 5470 observations. Significant similarity is observed.

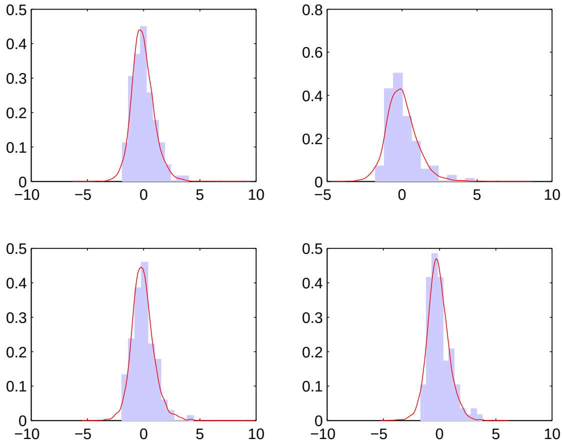

Figure S4: Monthly log precipitation data for July from 1895 to 2010 (116 observations) obtained from four stations in North Carolina (shown in histogram) are fitted using multivariate skewed-t distribution. Curves show the fitted distributions for the stations via maximum likelihood, using the **sn** R package.

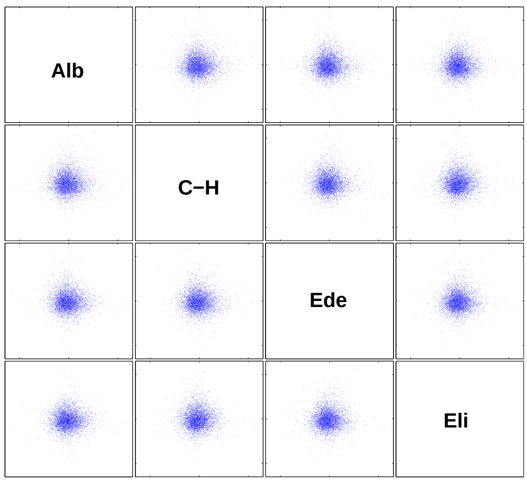

Figure S5: Covariance structure of monthly log precipitation when fitted with multivariate skewed-t distribution. Maximum likelihood estimation is used to obtain the fitted model using the **sn** R package.