Here $A = \{\mathbf{t} \in \mathbb{R}^{\mathbf{d}} : \mathbf{d_{GY}}(\mathbf{t}, \mathbf{F}, \mathbf{G}) \leq \mathbf{d_{GY}}(\zeta, \mathbf{F}, \mathbf{G})\}$, where ζ is any point in $\mathbb{R}^d$ such that $\Delta(\zeta, \mathbf{F}) = \eta$ and $\eta = G^{-1}(\alpha)$ is a large quantile of G. Then, we flag $\lfloor nd_n \rfloor$ observations. It is easy to see that,

$$\begin{aligned} d_n &= \sup_{\mathbf{t}\in\mathbf{A}}\{[1 - \hat{H}_n(\Delta(\mathbf{t}, \hat{\mathbf{F}}_{\mathbf{n}}))] - [\mathbf{1} - \mathbf{G}(\boldsymbol{\Delta}(\mathbf{t}, \hat{\mathbf{F}}_{\mathbf{n}}))]\}^+ \\ &= \sup_{\mathbf{t}\in\mathbf{A}}\{G(\Delta(\mathbf{t}, \hat{\mathbf{F}}_{\mathbf{n}})) - \hat{\mathbf{H}}_{\mathbf{n}}(\boldsymbol{\Delta}(\mathbf{t}, \hat{\mathbf{F}}_{\mathbf{n}}))\}^+ \\ &= \sup_{\Delta\geq\eta}\{G(\Delta) - \hat{H}_n(\Delta)\}^+ \end{aligned}$$

since d_{GY} is a non increasing function of the squared Mahalanobis distance of the point $\mathbf{t}$. We can rephrase Proposition 2. in **?**, that states the consistency property of the filter as follows. Consider a random vector $\mathbf{Y} = (\mathbf{X_1}, \ldots, \mathbf{X_d}) \sim \mathbf{F_0}$ and a pair of location and scatter estimators $\mathbf{T_{0n}}$ and $\mathbf{C_{0n}}$ such that $\mathbf{T_{0n}} \to \mu_{\mathbf{0}} = \mu(\mathbf{F_0}) \in \mathbb{R}^{\mathbf{d}}$ and $\mathbf{C_{0n}} \to \boldsymbol{\Sigma}_{\mathbf{0}} = \boldsymbol{\Sigma}(\mathbf{F_0})$ a.s.. Consider any continuous distribution function G and let $\hat{H}_n$ be the empirical distribution function of Δ_i and $H_0(t) = \Pr((\mathbf{Y} - \mu_{\mathbf{0}})^{\mathbf{t}}\boldsymbol{\Sigma}_{\mathbf{0}}^{-\mathbf{1}}(\mathbf{Y} - \mu_{\mathbf{0}}) \leq \mathbf{t})$. If the distribution G satisfies:

$$\max_{\mathbf{t}\in\mathbf{A}}\{d_{GY}(\mathbf{t}, \mathbf{F_0}, \mathbf{H_0}) - \mathbf{d_{GY}}(\mathbf{t}, \mathbf{F_0}, \mathbf{G})\} \leq \mathbf{0}, \tag{1}$$

where $A = \{\mathbf{t} \in \mathbb{R}^{\mathbf{d}} : \mathbf{d_{GY}}(\mathbf{t}, \mathbf{F_0}, \mathbf{G}) \leq \mathbf{d_{GY}}(\zeta, \mathbf{F_0}, \mathbf{G})\}$, where ζ is any point in $\mathbb{R}^d$ such that $\Delta(\zeta, \mathbf{F_0}) = \eta$ and $\eta = G^{-1}(\alpha)$ is a large quantile of G, then

$$\frac{n_0}{n} \to 0 \qquad \text{a.s.}$$

where

$$n_0 = \lfloor nd_n \rfloor.$$

Proof. Note that

$$d_{GY}(\mathbf{t}, \hat{\mathbf{F}}_{\mathbf{n}}, \hat{\mathbf{H}}_{\mathbf{n}}) - \mathbf{d_{GY}}(\mathbf{t}, \hat{\mathbf{F}}_{\mathbf{n}}, \mathbf{G}) = \mathbf{G}(\boldsymbol{\Delta}(\mathbf{t}, \mathbf{T_{0n}}, \mathbf{C_{0n}})) - \hat{\mathbf{H}}_{\mathbf{n}}(\boldsymbol{\Delta}(\mathbf{t}, \mathbf{T_{0n}}, \mathbf{C_{0n}}))$$

and condition in equation () is equivalent to

$$\max_{\Delta\geq\eta}\{G(\Delta) - H_0(\Delta)\} \leq 0,$$

The rest of the proof is the same as in Proposition 2. of **?**. □

Figure shows the bivariate scatter plot of WTS versus HTLD, HTLD versus WSBC and WSBC versus SUR where the GY-UBF and HS-UBF filters are applied, respectively. The bivariate observations with at least one component flagged as outlier are in blue, and outliers detected by the bivariate filter are in orange. We see that the HS-UBF identifies less outliers with respect to the GY-UBF.