Appendix A. Proofs

Proof. [Lemma 3.1] We can write:

$$\begin{aligned}\nabla_y \frac{\alpha f_1(y)}{f(y)} &= \frac{\int_{[\theta_\alpha,+\infty)} \nabla_y \log\varphi(y|\theta,\sigma^2)\varphi(y|\theta,\sigma^2)dG(\theta)}{\int_{(-\infty,+\infty)}\varphi(y|\theta,\sigma^2)dG(\theta)} \\ &- \frac{\int_{[\theta_\alpha,+\infty)}\varphi(y|\theta,\sigma^2)dG(\theta)}{\int_{(-\infty,+\infty))}\varphi(y|\theta,\sigma^2)dG(\theta)} \frac{\int_{(-\infty,+\infty)} \nabla_y \log\varphi(y|\theta,\sigma^2)\varphi(y|\theta,\sigma^2)dG(\theta)}{\int_{(-\infty,+\infty)}\varphi(y|\theta,\sigma^2)dG(\theta)} \\ &= \mathbb{E}\Big[\mathbb{1}\{\theta \geq \theta_\alpha\}\nabla_y \log\varphi(y|\theta,\sigma^2)|Y\Big] - \mathbb{E}\Big[\mathbb{1}\{\theta\geq\theta_\alpha\}|Y]\mathbb{E}[\nabla_y\log\varphi(y|\theta,\sigma^2)|Y\Big] \\ &= \text{Cov}\Big[\mathbb{1}\{\theta\geq\theta_\alpha\}, \nabla_y\log\varphi(y|\theta,\sigma^2)|Y\Big] \geq 0\end{aligned}$$

The last inequality holds since $\nabla_y \log\varphi(y|\theta,\sigma^2)$ is increasing in θ for each fixed σ^2, by the Gaussian assumption and the fact that covariance of monotone functions of θ is non-negative (see Schmidt (2014)) assuming the existence of $\mathbb{E}_{\theta|Y}[\nabla_y \log\varphi(Y|\theta,\sigma^2)|Y]$ which we assume. Nesting follows from the monotonicity of the $v_\alpha(y)$ criterion: monotonicity of $v_\alpha(y)$ implies that there exists t_α such that $\mathbb{1}\{v_\alpha(y)\geq \lambda_\alpha/(1+\lambda_\alpha)\} = \mathbb{1}\{y\geq t_\alpha\}$, hence if $\alpha_1 > \alpha_2$ and $\mathbb{P}(y\geq t_{\alpha_1}) = \alpha_1$ and $\mathbb{P}(y\geq t_{\alpha_2}) = \alpha_2$, then it must be that $t_{\alpha_1}\leq t_{\alpha_2}$, implying nestedness. ∎

Proof. [Lemma 3.2] Denote the three decision criteria, $v_1(y) = \mathbb{E}(\theta|Y=y)$, $v_2(y) = \mathbb{P}(\theta\geq G^{-1}(1-\alpha)|Y=y)$ and $v_3(y) = \mathbb{E}(\theta\mathbb{1}(\theta\geq G^{-1}(1-\alpha))|Y=y)$. Assuming that $\mathbb{E}[\theta|Y]<\infty$, $\mathbb{E}_{\theta|Y}[\nabla_y\log\varphi(y|\theta,\sigma^2)|Y]<\infty$ and $\mathbb{E}_{\theta|Y}[\theta\nabla_y\log\varphi(y|\theta,\sigma^2)|Y]<\infty$, the calculation leading to the proof of Lemma 1 shows:

$$\begin{aligned}\nabla_y v_1(y) &= \text{Cov}(\theta,\nabla_y\log\varphi(y|\theta)|Y=y)\\ \nabla_y v_2(y) &= \text{Cov}(\mathbb{1}(\theta\geq G^{-1}(1-\alpha)),\nabla_y\log\varphi(y|\theta)|Y=y)\\ \nabla_y v_3(y) &= \text{Cov}(\theta\mathbb{1}(\theta\geq G^{-1}(1-\alpha)),\nabla_y\log\varphi(y|\theta)|Y=y).\end{aligned}$$

Thus, the monotonicity of $\nabla_y\log\varphi(y|\theta,\sigma^2)$ implies they all yield identical rankings. ∎

Proof. [Proposition 3.3] The Bayes rule for the non-randomized selections can be characterized as,

$$\delta_i^* = \begin{cases}1, & \text{if } v_\alpha(y_i)\geq\tau_1^*(1-v_\alpha(y_i)-\gamma)+\tau_2^*\\ 0, & \text{if } v_\alpha(y_i)<\tau_1^*(1-v_\alpha(y_i)-\gamma)+\tau_2^*\end{cases}$$

with Karush-Kuhn-Tucker conditions,

$$\tau_1^*\Big(\mathbb{E}\Big[\sum_{i=1}^n\Big\{(1-v_\alpha(y_i))\delta_i^*-\gamma\delta_i^*\Big\}\Big]\Big) = 0 \tag{A.1}$$

$$\tau_2^*\Big(\mathbb{E}\Big[\sum_{i=1}^n\delta_i^*\Big]-\alpha n\Big) = 0 \tag{A.2}$$

$$\mathbb{E}\Big[\sum_{i=1}^n\Big\{(1-v_\alpha(y_i))\delta_i^*-\gamma\delta_i^*\Big\}\Big] \leq 0 \tag{A.3}$$

$$\frac{\mathbb{1}}{n}\mathbb{E}\Big[\sum_{i=1}^n\delta_i^*\Big]-\alpha \leq 0 \tag{A.4}$$

$$\tau_1^* \geq 0 \tag{A.5}$$

$$\tau_2^* \geq 0 \tag{A.6}$$

The Bayes rule takes the form of thresholding on the posterior tail probability $v_\alpha(y)$ and since $v_\alpha(y)$ is monotone in y as shown in Lemma 3.2, it is therefore a thresholding rule on Y, $\delta_i^* = \mathbb{1}\{y_i\geq t^*\}$ with cutoff t^* depending on the values of $(\tau_1^*,\tau_2^*,\alpha,\gamma)$. Condition (A.3) is equivalent to the condition that the marginal false discovery rate, mFDR, since it requires

$$\mathbb{E}\Big[\sum_{i=1}^{n}\{(1 - v_\alpha(y_i)\delta_i^*\}\Big]/\mathbb{E}\Big[\sum_{i=1}^{n}\delta_i^*\Big] \leq \gamma$$

and we can show that the left hand side quantity is precisely the mFDR since

$$\begin{aligned}
\text{mFDR}(t^*) =&\mathbb{P}(\delta_i^* = 1, \theta_i \leq \theta_\alpha)/\mathbb{P}(\delta_i^* = 1)\\
=&(1-\alpha)\int \mathbb{1}\{y \geq t^*\} f_0(y)dy/\int \mathbb{1}\{y \geq t^*\} f(y)dy\\
=&\int \mathbb{1}\{y \geq t^*\}(1 - v_\alpha(y))f(y)dy/\int \mathbb{1}\{y \geq t^*\} f(y)dy\\
=&\mathbb{E}\Big[\sum_{i=1}^{n}\{(1 - v_\alpha(y_i))\delta_i^*\}\Big]/\mathbb{E}\Big[\sum_{i=1}^{n}\delta_i^*\Big]\\
=&\frac{\int_{-\infty}^{\theta_\alpha}\tilde{\Phi}((t^* - \theta)/\sigma)dG(\theta)}{\int_{-\infty}^{+\infty}\tilde{\Phi}((t^* - \theta)/\sigma)dG(\theta)} \leq \gamma.
\end{aligned}$$

For any mixing distribution G, as t^* increases, it becomes less likely for condition (A.3) to bind. And as t^* approaches $-\infty$, left side of (A.3) approaches $1-\alpha-\gamma$, and hence we've restricted $\gamma < 1-\alpha$ to avoid cases where the condition (A.3) never binds. On the other hand, condition (A.4) is equivalent to

$$\mathbb{P}(\delta_i^* = 1) - \alpha = \int_{-\infty}^{+\infty}\tilde{\Phi}((t^* - \theta)/\sigma)dG(\theta) - \alpha \leq 0.$$

As t^* increases, it also becomes less likely that condition (A.4) binds. Therefore, we can define,

$$\begin{aligned}
t_1^* &= \min\{t : \text{mFDR}(t) - \gamma \leq 0\}\\
t_2^* &= \min\{t : \int_{-\infty}^{+\infty}\tilde{\Phi}((t - \theta)/\sigma)dG(\theta) - \alpha \leq 0\}
\end{aligned}$$

When $t_1^* < t_2^*$, the feasible region for Y defined by inequality (A.4) is a strict subset of that defined by inequality (A.3). When $t_1^* > t_2^*$, then the feasible region defined by inequality (A.3) is a strict subset of that defined by inequality (A.4). When $t_1^* = t_2^*$, the feasible regions coincide. This case occurs when $\text{mFDR}(t_1^*) = \gamma$ and $\mathbb{P}(y \geq t_2^*) = \alpha$, so $\mathbb{E}[v_\alpha(Y)\mathbb{1}\{v_\alpha(Y) \geq \lambda^*\}] = \alpha - \alpha\gamma$ with $v_\alpha(t_1^*) = \lambda^*(\alpha, \gamma)$. Again, the strict thresholding enforced by the statement of the proposition can be relaxed slightly by randomizing the selection probability of the last unit so that the active constraint is satisfied exactly. ■

Proof. [Proposition 4.1] In the proof we will suppress the dependence of λ^* on the (α, γ). The argument is very similar to the proof for Proposition 3.3, except that now the feasible region defined by constraint (A.3) and (A.4) is a two-dimensional region for (y_i, σ_i). Since the posterior tail probability $v_\alpha(y, \sigma)$ is monotone in y for any fixed σ as a result of Lemma 3.2, the optimal rule can be reformulated as a thresholding rule on Y again, $\delta_i^* = \mathbb{1}\{y_i > t_\alpha(\lambda^*, \sigma_i)\}$ except now the threshold value also depends on σ_i.

Now consider the constraint (A.3) and (A.4). Condition (A.3) is equivalent to the condition that,

$$\frac{\mathbb{E}\Big[\sum_{i=1}^{n}\{(1 - v_\alpha(y_i, \sigma_i))\delta_i^*\}\Big]}{\mathbb{E}\Big[\sum_{i=1}^{n}\delta_i^*\Big]} - \gamma = \frac{\int\int_{-\infty}^{\theta_\alpha}\tilde{\Phi}((t_\alpha(\lambda^*, \sigma) - \theta)/\sigma)dG(\theta)dH(\sigma)}{\int\int_{-\infty}^{+\infty}\tilde{\Phi}((t_\alpha(\lambda^*, \sigma) - \theta)/\sigma)dG(\theta)dH(\sigma)} - \gamma \leq 0.$$

For any marginal distributions G and H of (θ, σ) and for a fixed pair of (α, γ), as λ^* increases, $t_\alpha(\lambda^*, \sigma)$ also increases for any $\sigma > 0$ and therefore it is less likely for condition (A.3) to bind. On the other hand, condition (A.4) is equivalent to,

$$\mathbb{P}(\delta_i^* = 1) - \alpha = \int\int \tilde{\Phi}((t_\alpha(\lambda^*, \sigma) - \theta)/\sigma)dG(\theta)dH(\sigma) - \alpha \leq 0.$$

So as λ^* increases, it is also less likely for condition (A.4) to bind. Thus, when $\lambda_1^* < \lambda_2^*$, the feasible region on (Y, σ) defined by inequality constraint (A.4) is a strict subset of that defined by inequality (A.3). When $\lambda_1^* > \lambda_2^*$, then the feasible region defined by inequality (A.3) is a strict subset of that defined by inequality (A.4). When $\lambda_1^* = \lambda_2^*$, the feasible regions coincide; this case occurs when

$$\mathbb{E}\Big[v_\alpha(Y,\sigma)\mathbb{1}\{v_\alpha(Y,\sigma) \geq \lambda^*\}\Big] = \alpha - \alpha\gamma$$

where the expectation is taken with respect to the joint distribution of (Y, σ).

Finally regarding the existence of λ^*, note that existence of a solution for λ_2^* for any $\alpha \in (0,1)$ follows from the fact that for any fixed α, $f_2(\alpha,\lambda) = \mathbb{P}(v_\alpha(y,\sigma) > \lambda) - \alpha$, is a decreasing function in λ and $\lambda_2^*(\alpha)$ is defined as the zero-crossing point of $f_2(\alpha,\lambda)$. Note that $f_2(\alpha, 0) = 1-\alpha$ and $f_2(\alpha,1) = -\alpha$. Therefore for any $\alpha \in (0,1)$, we can always find a $\lambda_2^*(\alpha) \in (0,1)$ such that $f_2(\alpha, \lambda_2^*(\alpha)) = 0$. Now consider $f_1(\alpha,\gamma,\lambda) = \mathbb{E}[(1 - v_\alpha(y,\sigma) - \gamma)\mathbb{1}\{v_\alpha(y,\sigma) > \lambda\}]$. For a fixed pair of (α,γ), $\lambda_1^*(\alpha,\gamma)$ is defined as the zero crossing point of $f_1(\alpha,\gamma,\lambda)$. Note that $f_1(\alpha,\gamma,\lambda)$ decreases first and then increases in λ with its minimum achieved at $\lambda = 1-\gamma$. We also know that $f_1(\alpha,\gamma,0) = 1-\gamma-\mathbb{E}[v_\alpha(y,\sigma)] = 1-\gamma-\alpha$ and $f_1(\alpha,\gamma,1) = 0$. Hence as long as $\gamma < 1-\alpha$, the zero-crossing $\lambda_1^*(\alpha,\gamma)$ exists. The condition $\gamma < 1-\alpha$ is imposed to rule out cases where FDR constraint never binds. ■

Proof. [Lemma 4.2] Note that for any cutoff value λ, the mFDR can be expressed as,

$$\text{mFDR}(\alpha,\lambda) = \mathbb{E}[(1 - v_\alpha(y_i,\sigma_i))\mathbb{1}\{v_\alpha(y_i,\sigma_i) \geq \lambda\}]/\mathbb{P}[v_\alpha(y_i,\sigma_i) \geq \lambda].$$

Thus, mFDR depends both on the cutoff value λ and on α since v_α, is a function of α, and consequently its density function is also indexed by α.

First, we will show that ∇_λ mFDR$(\alpha,\lambda) \leq 0$ for all $\alpha \in (0,1)$. Differentiating with respect to λ gives,

$$\begin{aligned}
\nabla_\lambda \frac{\int_\lambda^1 (1-v) f_v(v;\alpha)dv}{\int_\lambda^1 f_v(v;\alpha)dv} &= \frac{-(1-\lambda)f_v(\lambda;\alpha)\int_\lambda^1 f_v(v;\alpha)dv + \int_\lambda^1(1-v)f_v(v;\alpha)dv f_v(\lambda;\alpha)}{(\int_\lambda^1 f_v(v;\alpha)dv)^2} \\
&= \frac{f_v(\lambda;\alpha)}{(\int_\lambda^1 f_v(v;\alpha)dv)^2}\Big(\int_\lambda^1 (1-v)f_v(v;\alpha)dv - \int_\lambda^1 (1-\lambda)f_v(v;\alpha)dv\Big) \\
&\leq 0.
\end{aligned}$$

Next, to establish that ∇_α mFDR$(\alpha,\lambda) \leq 0$, we differentiate with respect to α, to obtain,

$$\begin{aligned}
&\nabla_\alpha \frac{\int_\lambda^1 (1-v) f_v(v;\alpha)dv}{\int_\lambda^1 f_v(v;\alpha)dv} \\
&= \frac{\int_\lambda^1 (1-v)\nabla_\alpha \log f_v(v;\alpha) f_v(v;\alpha)dv}{\int_\lambda^1 f_v(v;\alpha)dv} - \frac{\int_\lambda^1 (1-v)f_v(v;\alpha)dv}{\int_\lambda^1 f_v(v;\alpha)dv}\frac{\int_\lambda^1 \nabla_\alpha \log f_v(v;\alpha) f_v(v;\alpha)dv}{\int_\lambda^1 f_v(v;\alpha)dv} \\
&= \frac{\mathbb{E}[(1-v)\nabla_\alpha \log f_v(v;\alpha)\mathbb{1}\{v \geq \lambda\}]}{\mathbb{P}(v \geq \lambda)} - \frac{\mathbb{E}[(1-v)\mathbb{1}\{v \geq \lambda\}]}{\mathbb{P}(v \geq \lambda)}\frac{\mathbb{E}[\nabla_\alpha \log f_v(v;\alpha)\mathbb{1}\{v \geq \lambda\}]}{\mathbb{P}(v \geq \lambda)} \\
&= \mathbb{E}[(1-v)\nabla_\alpha \log f_v(v;\alpha) | v \geq \lambda] - \mathbb{E}[1-v | v \geq \lambda]\mathbb{E}[\nabla_\alpha \log f_v(v;\alpha) | v \geq \lambda] \\
&= \text{cov}[1-v, \nabla_\alpha \log f_v(v;\alpha) | v \geq \lambda] \leq 0.
\end{aligned}$$

where the last inequality holds because $\nabla_\alpha \log f_v(v;\alpha)$ is non-decreasing in v.

Now suppose we have the cutoff value $\lambda_1^*(\alpha_2,\gamma)$ such that,

$$\mathbb{E}[(1 - v_{\alpha_2}(y_i,\sigma_i))\mathbb{1}\{v_{\alpha_2}(y_i,\sigma_i) > \lambda_1^*(\alpha_2,\gamma)\}]/\mathbb{P}(v_{\alpha_2}(y_i,\sigma_i) > \lambda_1^*(\alpha_2,\gamma)] = \gamma.$$

If we maintain the same cutoff value for $v_{\alpha_1}(y_i,\sigma_i)$ with $\alpha_1 > \alpha_2$, given the second property of mFDR, we know,

$$\mathbb{E}[(1 - v_{\alpha_1}(y_i,\sigma_i))\mathbb{1}\{v_{\alpha_1}(y_i,\sigma_i) \geq \lambda_1^*(\alpha_2,\gamma)\}]/\mathbb{P}(v_{\alpha_1}(y_i,\sigma_i) \geq \lambda_1^*(\alpha_2,\gamma)) \leq \gamma.$$

If equality holds, then by definition we have $\lambda_1^*(\alpha_2,\gamma) = \lambda_1^*(\alpha_1,\gamma)$, if strict inequality holds, then by the first property of mFDR, in order to increase mFDR level to be equal to γ, we must have $\lambda_1^*(\alpha_1,\gamma) < \lambda_1^*(\alpha_2,\gamma)$. ■

Proof. [Corollary 4.3] For any $\alpha_1 > \alpha_2$, we have $v_{\alpha_1}(y,\sigma) \geq v_{\alpha_2}(y,\sigma)$ for all pair of $(y,\sigma) \in \mathbb{R}\times\mathbb{R}_+$. When the condition in Lemma 4.2 holds, then $v_{\alpha_1}(y,\sigma) \geq v_{\alpha_2}(y,\sigma) > \lambda_1^*(\alpha_2,\gamma) \geq \lambda_1^*(\alpha_1,\gamma)$, which implies $\Omega^{FDR}_{\alpha_2,\gamma} \subseteq \Omega^{FDR}_{\alpha_1,\gamma}$. ■

Proof. [Lemma 4.4] Since $t_\alpha(\lambda_2^*(\alpha),\sigma)$ defines the boundary of the selection region under the capacity constraint for a fixed level α. The condition imposed that as α increases, for each fixed σ, the thresholding value for Y decreases, hence nestedness of the selection region. ■

Proof. [Lemma 4.5] Based on results in Lemma 4.2 and Lemma 4.4 and the fact that $\Omega_{\alpha,\gamma} = \Omega^{FDR}_{\alpha,\gamma} \cap \Omega^C_\alpha$, we have nestedness of the selection set. ■

Proof. [Proposition C.1] The capacity constraint requires that

$$\alpha = \mathbb{P}(M(y,\sigma) \geq C_2^*(\alpha)) = \int\int \mathbb{1}\{M(y,\sigma) \geq C_2^*(\alpha)\} f(y|\theta,\sigma)dG(\theta)dH(\sigma)$$

For any $\alpha_1 > \alpha_2$, it is then clear that $C_2^*(\alpha_1) \leq C_2^*(\alpha_2)$. Given the monotonity of $M(y,\sigma)$ for each fixed σ established in Lemma 3.2, the selection set based on capacity constraint is nested. For FDR constraint, we require

$$\gamma = \frac{\int\int_{-\infty}^{\theta_\alpha} \mathbb{1}\{M(y,\sigma) \geq C_1^*(\alpha,\gamma)\} f(y|\theta,\sigma)dG(\theta)dH(\sigma)}{\int\int \mathbb{1}\{M(y,\sigma) \geq C_1^*(\alpha,\gamma)\} f(y|\theta,\sigma)dG(\theta)dH(\sigma)} \tag{A.7}$$

Fix γ, it suffices to show that if $\alpha_1 > \alpha_2$, then $C_1^*(\alpha_1,\gamma) \leq C_1^*(\alpha_2,\gamma)$. First, solve for $C_1^*(\alpha_2,\gamma)$ from equation (A.7). Now suppose we use this same thresholding value when we increase capacity to $\alpha_1 > \alpha_2$, we evaluate the right hand side of equation (A.7). Since $\theta_{\alpha_1} \leq \theta_{\alpha_2}$, then the numerator decreases,

$$\begin{aligned}&\int\int_{-\infty}^{\theta_{\alpha_1}} \mathbb{1}\{M(y,\sigma) \geq C_1^*(\alpha_2,\gamma)\} f(y|\theta,\sigma)dG(\theta)dH(\sigma)\\ \leq &\int\int_{-\infty}^{\theta_{\alpha_2}} \mathbb{1}\{M(y,\sigma) \geq C_1^*(\alpha_2,\gamma)\} f(y|\theta,\sigma)dG(\theta)dH(\sigma),\end{aligned}$$

while the denominator does not change. The only way to satisfy the equality (A.7) again is to decrease the thresholding value, therefore $C_1^*(\alpha_1,\gamma) \leq C_1^*(\alpha_2,\gamma)$. The result in the Proposition is then reached since the selection set is the intersection of the selection set under capacity constraint and that under FDR constraint. ■

Proof. [Lemma 5.1] The logarithm of the Gamma density of S_i takes the form,

$$\log\Gamma(S_i|r_i,\sigma_i^2) = r_i\log(r_i/\sigma_i^2) - \log(\Gamma(r_i)) + (r_i-1)\log S_i - S_i\frac{r_i}{\sigma_i^2},$$

hence $\nabla_s\Gamma(s|r,\sigma^2) = \Gamma(s|r,\sigma^2)\Big(\frac{r-1}{s} - \frac{r}{\sigma^2}\Big)$. Fixing y and differentiating with respect to s, we have,

$$\begin{aligned}\nabla_s v_\alpha(y,s) =& \frac{\int\int_{\theta_\alpha}^{+\infty} f(y|\theta,\sigma^2)\Gamma(s|r,\sigma^2)\Big[\frac{r-1}{s} - \frac{r}{\sigma^2}\Big]dG(\theta,\sigma^2)}{\int\int f(y|\theta,\sigma^2)\Gamma(s|r,\sigma^2)dG(\theta,\sigma^2)}\\ &- \frac{\int\int_{\theta_\alpha}^{+\infty} f(y|\theta,\sigma^2)\Gamma(s|r,\sigma^2)dG(\theta,\sigma^2)}{\int\int f(y|\theta,\sigma^2)\Gamma(s|r,\sigma^2)dG(\theta,\sigma^2)}\frac{\int\int f(y|\theta,\sigma^2)\Gamma(s|r,\sigma^2)\Big[\frac{r-1}{s} - \frac{r}{\sigma^2}\Big]dG(\theta,\sigma^2)}{\int\int f(y|\theta,\sigma^2)\Gamma(s|r,\sigma^2)dG(\theta,\sigma^2)}\\ =& \mathbb{E}\Big[\mathbb{1}\{\theta \geq \theta_\alpha\}\Big(\frac{r-1}{s} - \frac{r}{\sigma^2}\Big)|Y=y,S=s\Big] - \mathbb{E}\Big[\mathbb{1}\{\theta \geq \theta_\alpha\}|Y=y,S=s\Big]\mathbb{E}\Big[\frac{r-1}{s} - \frac{r}{\sigma^2}|Y=y,S=s\Big]\\ =& -\text{Cov}\Big[\mathbb{1}\{\theta \geq \theta_\alpha\}, \frac{r}{\sigma^2}|Y=y,S=s\Big]\end{aligned}$$

The covariance term can take either sign since we do not restrict the distribution G, so $v_\alpha(Y,S)$ need not be monotone in S. On the other hand, if we fix s and differentiate with respect to y,

$$\begin{aligned}\nabla_y v_\alpha(y,s) &= \mathbb{E}\Big[\mathbb{1}\{\theta \geq \theta_\alpha\}\Big[-\frac{y-\theta}{\sigma^2/T}\Big]|Y=y,S=s\Big] - \mathbb{E}\Big[\mathbb{1}\{\theta\geq\theta_\alpha\}|Y=y,S=s\Big]\mathbb{E}\Big[-\frac{y-\theta}{\sigma^2/T}|Y=y,S=s\Big]\\ &= \mathrm{Cov}\Big[\mathbb{1}\{\theta\geq\theta_\alpha\},\frac{\theta-y}{\sigma^2/T}|Y=y,S=s\Big].\end{aligned}$$

Again, the covariance term can take either sign, depending on the correlation of θ and σ^2 conditional on (Y,S). Therefore, fixing S, $v_\alpha(Y,S)$ need not be a monotone function of Y. ■

Proof. [Proposition 5.2] The proof is very similar to that of Proposition 4.1, the only difference is that we can no longer formulate the decision rule by simply thresholding on Y because the transformation $v_\alpha(Y,S)$ need not be monotone in Y for fixed values of S as shown in Lemma 5.1, hence $\lambda_1^*(\alpha,\gamma)$ and $\lambda_2^*(\alpha)$ must now be defined directly through the random variable v_α. The first constraint states that

$$\mathbb{E}\Big[\sum_{i=1}^n\{(1-v_\alpha(y_i,s_i))\delta_i^*\Big]/\mathbb{E}\Big[\sum_{i=1}^n\delta_i^*\Big]\leq\gamma$$

with $\delta_i^* = 1\{v_\alpha(y_i,s_i)\geq\lambda^*\}$ For each fixed α, let the density function for $v_\alpha(Y,S)$ be denoted as $f_v(\cdot;\alpha)$, then the constraints can be formulated as

$$\int_{\lambda^*}^1(1-v)f_v(v;\alpha)dv/\int_{\lambda^*}^1 f_v(v;\alpha)dv$$

which is non-increasing in λ^*, hence the constraint becomes less likely to bind as λ^* increases. On the other hand the second constraint states that,

$$\mathbb{P}(\delta_i^*=1)-\alpha=\int_{\lambda^*}^1 f_v(v;\alpha)dv-\alpha$$

For each fixed $\alpha\in(0,1)$, this constraint also becomes less likely to bind as λ^* increases. ■

Proof. [Theorem 6.1] To prove Theorem 6.1, we first introduce some additional notation and prove several lemmas. Let

$$\begin{aligned}H_{n,0}(t) &= 1-H_n(t)=\frac{1}{n}\sum_{i=1}^n 1\{v_{\alpha,i}\geq t\}\\ H_{n,1}(t) &= \frac{1}{n}\sum_{i=1}^n(1-v_{\alpha,i})1\{v_{\alpha,i}\geq t\}\\ Q_n(t) &= H_{n,1}(t)/H_{n,0}(t)\\ V_n(t) &= \frac{1}{n}\sum_{i=1}^n 1\{v_{\alpha,i}\geq t\}1\{\theta_i\leq\theta_\alpha\}\\ H_0(t) &= 1-H(t)=\mathbb{P}(v_{\alpha,i}\geq t)\\ H_1(t) &= \mathbb{E}\Big[(1-v_{\alpha,i})1\{v_{\alpha,i}\geq t\}\Big]\\ Q(t) &= H_1(t)/H_0(t)\end{aligned}$$

Lemma A.1. *Under Assumption 1, as $n\to\infty$,*

$$\sup_{t\in[0,1]}|H_{n,0}(t)-H_0(t)|\xrightarrow{p}0$$

$$\sup_{t\in[0,1]}|H_{n,1}(t)-H_1(t)|\xrightarrow{p}0$$

Proof. [Proof of Lemma A.1] Under Assumption 1 and the fact that $v_{\alpha,i} \in [0,1]$, the weak law of large numbers implies that we have for any $t \in [0,1]$, as $n \to \infty$,

$$H_{n,0}(t) \xrightarrow{p} H_0(t)$$
$$H_{n,1}(t) \xrightarrow{p} H_1(t)$$

By the Glivenko-Cantalli theorem, the first result is immediate. To prove the second result, it suffices to show that for any $\epsilon > 0$, as $n \to \infty$,

$$\mathbb{P}\Big(\sup_{t\in[0,1]}|H_{n,1}(t) - H_1(t)| > \epsilon\Big) \to 0.$$

It is clear that since $v_{\alpha,i}$ has a continuous distribution, $H_1(t)$ is a monotonically decreasing and bounded function in t with $H_1(0) = 1-\alpha$ and $H_1(1) = 0$. It is also clear that the function $H_{n,1}(t)$ is monotonically decreasing in t, so we can find $m_\epsilon < \infty$ points such that $0 = t_0 < t_1 < \cdots < t_{m_\epsilon} = 1$, and for any $j \in \{1, 2, \ldots, m_\epsilon\}$, we have $H_1(x_j) - H_1(x_{j-1}) \le \epsilon/2$. For any $t \in [0,1]$, there exists a j such that $x_{j-1} \le t \le x_j$ and

$$\begin{aligned}
H_{n,1}(t) - H_1(t) &\le H_{n,1}(t_{j-1}) - H_1(t_j) \\
&= (H_{n,1}(t_{j-1}) - H_1(t_{j-1})) + (H_1(t_{j-1}) - H_1(t_j)) \\
&\le |H_{n,1}(x_{j-1}) - H_1(x_{j-1})| + \epsilon/2 \le \max_j |H_{n,1}(t_{j-1}) - H_1(t_{j-1})| + \epsilon/2
\end{aligned}$$

Likewise we can show that $H_{n,1}(t) - H_1(t) \ge -\max_j |H_{n,1}(t_j) - H_1(t_j)| - \epsilon/2$, hence

$$\sup_{t\in[0,1]} |H_{n,1}(t) - H_1(t)| \le \max_j |H_{n,1}(t_j) - H_1(t_j)| + \epsilon/2.$$

Since m_ϵ is finite then for any $\delta > 0$, there exists N such that for all $n \ge N$,

$$\mathbb{P}\Big(\max_j |H_{n,1}(t_j) - H_1(t_j)| \ge \epsilon/2\Big) \le \delta$$

which then implies that

$$\begin{aligned}
\mathbb{P}\Big(\sup_{t\in[0,1]} |H_{n,1}(t) - H_1(t)| \ge \epsilon\Big) &\le \mathbb{P}\Big(\frac{\epsilon}{2} + \max_j |H_{n,1}(t_j) - H_1(t_j)| \ge \epsilon\Big) \\
&= \mathbb{P}\Big(\max_j |H_{n,1}(t_j) - H_1(t_j)| \ge \epsilon/2\Big) \to 0
\end{aligned}$$

■

Lemma A.2. *Under Assumption 1 and $\alpha < 1 - \gamma$, $Q(1-\gamma) < \gamma$.*

Proof. [Proof of Lemma A.2] Define $\bar{Q}(t) = \mathbb{E}\Big[(1 - v_{\alpha,i} - \gamma)1\{v_{\alpha,i} \ge t\}\Big]$, then $Q(t) = \gamma$ implies $\bar{Q}(t) = 0$. Since $Q(t)$ is monotonically decreasing in t as shown in the proof of Lemma 4.2 , it suffices to prove that $\bar{Q}(1-\gamma) < 0$. To this end, note that $\nabla_t \bar{Q}(t) < 0$ for $t < 1-\gamma$ and $\nabla_t \bar{Q}(t) > 0$ for $t > 1-\gamma$, hence $\bar{Q}(t)$ obtains its minimun value at $t = 1-\gamma$. Note that $\bar{Q}(0) = 1-\gamma-\alpha$ and $\bar{Q}(1) = 0$, thus $\bar{Q}(1-\gamma) < 0$. ■

Now we are ready to prove Theorem 6.1. We prove the first statement, as the second statement can be shown with a similar argument. First we show that, $\sup_{t\le 1-\gamma}\Big|Q_n(t) - Q(t)\Big| \xrightarrow{p} 0$, since,

$$\begin{aligned}
\Big|Q_n(t) - Q(t)\Big| &= \Big|\frac{H_0(t)H_{n,1}(t) - H_1(t)H_{n,0}(t)}{H_{n,0}(t)H_0(t)}\Big| = \Big|\frac{H_0(t)(H_{n,1}(t) - H_1(t)) - H_1(t)(H_{n,0}(t) - H_0(t))}{H_{n,0}(t)H_0(t)}\Big| \\
&\le \frac{H_0(0)\sup_t|H_{n,1}(t) - H_1(t)| + H_1(0)\sup_t|H_{n,0}(t) - H_0(t)|}{H_0(1-\gamma)\Big(H_0(1-\gamma) - \sup_t|H_{n,0}(t) - H_0(t)|\Big)} \xrightarrow{p} 0
\end{aligned}$$

uniformly for any $t \leq 1-\gamma$. The last inequality holds because $\min_{t\leq 1-\gamma} H_0(t) = H_0(1-\gamma)$ by monotonicity of $H_0(t)$. With a similar argument, we can also show that $\sup_{t\leq 1-\gamma}\left|\frac{V_n(t)}{H_{n,0}(t)} - Q(t)\right| \xrightarrow{p} 0$. Using this result and the fact that $Q(1-\gamma) < \gamma$ by Lemma A.2, we have $\mathbb{P}\Big(|Q_n(1-\gamma) - Q(1-\gamma)| < \frac{\gamma - Q(1-\gamma)}{2}\Big) \to 1$ and therefore, $\mathbb{P}(Q_n(1-\gamma) < \gamma) \to 1$ and $\mathbb{P}(\lambda_{2n} \leq 1-\gamma) \to 1$ by the definition of λ_{2n}. Since $\lambda_n \leq \lambda_{2n}$ by definition, we also have $\mathbb{P}(\lambda_n \leq 1-\gamma) \to 1$. On the other hand,

$$Q_n(\lambda_n) - \frac{V_n(\lambda_n)}{H_{n,0}(\lambda_n)} \geq \inf_{t\leq 1-\gamma}\Big(Q_n(t) - Q(t) + Q(t) - V_n(t)/H_{n,0}(t)\Big) = o_p(1)$$

Since $Q_n(\lambda_n) \leq Q_n(\lambda_{2n}) \leq \gamma$, it follows that

$$\frac{V_n(\lambda_n)}{H_{n,0}(\lambda_n)\bigvee 1} \leq \frac{V_n(\lambda_n)}{H_{n,0}(\lambda_n)} \leq \gamma + o_p(1).$$

Since $\frac{V_n(\lambda_n)}{H_{n,0}(\lambda_n)\bigvee 1}$ is upper bounded by 1, by Fatou's lemma, we have

$$\limsup_{n\to\infty}\mathbb{E}\Big[\frac{V_n(\lambda_n)}{H_{n,0}(\lambda_n)\bigvee 1}\Big] \leq \gamma$$

■

Lemma A.3. *Under Assumption 1 and 2, as $n \to \infty$,*

$$\hat{\theta}_\alpha \to \theta_\alpha \quad a.s.$$

Proof. [Proof of Lemma A.3] See Lemma 21.2 in van der Vaart (2000). ■

Lemma A.4. *Under Assumption 1 and 2, as $n \to \infty$,*

$$\sup_i |\hat{v}_{\alpha,i} - v_{\alpha,i}| \to 0 \quad a.s.$$

Proof. [Proof of Lemma A.3] Since

$$\hat{v}_{\alpha,i} = \frac{\int_{\hat{\theta}_\alpha}^{+\infty} f(D_i|\theta)d\hat{G}_n(\theta)}{\int_{-\infty}^{+\infty} f(D_i|\theta)d\hat{G}_n(\theta)}$$

where we denote D_i as data with a density function $f(D_i|\theta)$. When variances are known, then $D_i = \{y_i, \sigma_i\}$ and $f(D_i|\theta) = \frac{1}{\sigma_i}\varphi((y_i-\theta)/\sigma_i)$ and when variances are unknown, then $D_i = \{y_i, s_i\}$ and $f(D_i|\theta) = \frac{1}{\sqrt{\sigma^2/T}}\varphi((y_i-\theta)/\sqrt{\sigma^2/T})\Gamma(s_i|r, \sigma^2/r)$ with $r = (T-1)/2$ with $\varphi(\cdot)$ and $\Gamma(\cdot|\cdot,\cdot)$ being the standard normal and gamma density function, respectively.

We first analyze the denominator and prove

$$\sup_x\left|\int_{-\infty}^{+\infty} f(x|\theta)d\hat{G}_n(\theta) - \int_{-\infty}^{+\infty} f(x|\theta)dG(\theta)\right| \to 0 \quad a.s. \tag{A.8}$$

Let $f_n(x) = \int f(x|\theta)d\hat{G}_n(\theta)$ and $f(x) = \int f(x|\theta)dG(\theta)$. Under Assumption 2, we have $\frac{1}{2}\int\Big(\sqrt{f_n(x)} - \sqrt{f(x)}\Big)^2 d\mu(x) \to 0$ almost surely, which implies that $\int|f_n(x) - f(x)|dx \to 0$ almost surely. If $f_n(x)$ and $f(x)$ are Lipschitz continuous, we proceed by contradiction. Suppose (A.8) doesn't hold, then there exists $\epsilon > 0$ and a sequence $\{x_n\}_{n\geq 1}$ such that $|f_n(x_n) - f(x_n)| \geq \epsilon$ for all n. By Lipschitz continuity of f_n and f, there exists C such that

$$|f_n(x_n+\delta) - f_n(x_n)| \leq C\|\delta\|$$
$$|f(x_n+\delta) - f(x_n)| \leq C\|\delta\|$$

And therefore there exists $\eta > 0$ and for all $\|y - x_n\| \leq \eta$, $|f_n(y) - f(y)| \geq \epsilon/2$, which then implies

$$\int|f_n(x) - f(x)|dx \geq \int 1\{\|y-x_n\| \leq \eta\}|f_n(y) - f(y)|dy \geq \frac{\epsilon}{2}\int 1\{\|y - x_n\| \leq \eta\}dy$$

which contradicts $\int |f_n(x) - f(x)|dx \to 0$ almost surely. To prove that the functions f_n and f are Lipschitz continuous. Note that it suffices to prove that for each fixed parameter θ, $|f(x|\theta)-f(y|\theta)| \leq C_\theta \|x-y\|$ and $\sup_\theta C_\theta < \infty$. This clearly holds for the Gaussian density since the Gaussian density is everywhere differentiable and has bounded first derivative. Under Assumption 1 with $T \geq 4$, the Gamma density is also everywhere differentiable and has bounded first derivative, and thus is Lipschitz continuous.

We next analyze the numerator and show

$$\sup_x \left| \int_{\hat{\theta}_\alpha}^{+\infty} f(x|\theta) d\hat{G}_n(\theta) - \int_{\theta_\alpha}^{+\infty} f(x|\theta) dG(\theta) \right| \to 0 \quad a.s. \tag{A.9}$$

Note that

$$\begin{aligned}
&\left| \int_{\hat{\theta}_\alpha}^{+\infty} f(x|\theta) d\hat{G}_n(\theta) - \int_{\theta_\alpha}^{+\infty} f(x|\theta) dG(\theta) \right| \\
&\leq \left| \int_{\hat{\theta}_\alpha}^{+\infty} f(x|\theta) d\hat{G}_n(\theta) - \int_{\theta_\alpha}^{+\infty} f(x|\theta) d\hat{G}_n(\theta) \right| + \left| \int_{\theta_\alpha}^{+\infty} f(x|\theta) d\hat{G}_n(\theta) - \int_{\theta_\alpha}^{+\infty} f(x|\theta) dG(\theta) \right|.
\end{aligned}$$

The first term converges to 0 uniformly due to Lemma A.3. To show the second term also converges to zero uniformly, we make use of the result that if $\hat{G}_n$ weakly converges to G, which holds under Assumption 2, then $\sup_{g \in \mathcal{BL}} |\int g d\hat{G}_n - \int g dG| \to 0$ where $\mathcal{BL}$ is the class of bounded Lipschitz continuous functions. Note that $f(x|\theta)1\{\theta \geq \theta_\alpha\}$ is bounded and continuous except at $\theta = \theta_\alpha$. So we construct a smoothed version of $f(x|\theta)1\{\theta \geq \theta_\alpha\}$, denoted as $g(x|\theta)$, by replacing $1\{\theta \geq \theta_\alpha\}$ by a piecewise linear function taking value zero for $\theta < \theta_\alpha$ and value 1 for $\theta \geq \theta_\alpha + \epsilon$ and taking the form $-\theta_\alpha/\epsilon + \theta/\epsilon$ for $\theta \in [\theta_\alpha, \theta_\alpha + \epsilon]$, then $g \in \mathcal{BL}$. The result (A.9) then holds by showing that

$$\sup_x \left| \int_{\theta_\alpha}^{\theta_\alpha+\epsilon} f(x|\theta) d\hat{G}_n(\theta) + \int_{\theta_\alpha}^{\theta_\alpha+\epsilon} f(x|\theta) dG(\theta) \right| \to 0 \quad a.s.$$

which holds by Assumptions 1 and 2. ■

Proof. [Proof of Theorem 6.2] Define analogously

$$\begin{aligned}
\hat{H}_{n,0}(t) &= \frac{1}{n} \sum_{i=1}^n 1\{\hat{v}_{\alpha_i} \geq t\} \\
\hat{H}_{n,1}(t) &= \frac{1}{n} \sum_{i=1}^n (1 - \hat{v}_{\alpha,i}) 1\{\hat{v}_{\alpha,i} \geq t\} \\
\hat{Q}_n(t) &= \hat{H}_{n,1}(t) / \hat{H}_{n,0}(t)
\end{aligned}$$

We first show

$$\begin{aligned}
\sup_{t \in [0,1]} |\hat{H}_{n,0}(t) - H_0(t)| \xrightarrow{p} 0 \\
\sup_{t \in [0,1]} |\hat{H}_{n,1}(t) - H_1(t)| \xrightarrow{p} 0
\end{aligned}$$

We will prove the second statement now, and the first can be proved using a similar argument. To prove the second statement, it suffices to show that

$$\sup_{t \in [0,1]} \left| \hat{H}_{n,1}(t) - H_{n,1}(t) \right| \xrightarrow{p} 0$$

To this end, note

$$\begin{aligned}
&\sup_{t\in[0,1]}\left|\frac{1}{n}\sum_i(1-\hat{v}_{\alpha,i})1\{\hat{v}_{\alpha_i}\geq t\}-\frac{1}{n}\sum_i(1-v_{\alpha,i})1\{v_{\alpha,i}\geq t\}\right|\\
&=\sup_{t\in[0,1]}\left|\frac{1}{n}\sum_i(1-\hat{v}_{\alpha,i})1\{\hat{v}_{\alpha_i}\geq t\}-\frac{1}{n}\sum_i(1-v_{\alpha,i})1\{\hat{v}_{\alpha,i}\geq t\}\right|\\
&+\sup_{t\in[0,1]}\left|\frac{1}{n}\sum_i(1-v_{\alpha,i})1\{\hat{v}_{\alpha,i}\geq t\}-\frac{1}{n}\sum_i(1-v_{\alpha,i})1\{v_{\alpha,i}\geq t\}\right|\\
&\leq\frac{1}{n}\sum_i|\hat{v}_{\alpha,i}-v_{\alpha,i}|+\sup_{t\in[0,1]}\frac{1}{n}\sum_i|1\{\hat{v}_{\alpha,i}\geq t\}-1\{v_{\alpha,i}\geq t\}|
\end{aligned}$$

The first term is implied by the result in Lemma A.4. The second term can be written as

$$\begin{aligned}
&\sup_{t\in[0,1]}\frac{1}{n}\sum_i\left|1\{\hat{v}_{\alpha,i}\geq t\}-1\{v_{\alpha,i}\geq t\}\right|\\
&=\sup_{t\in[0,1]}\frac{1}{n}\sum_i\Big[1\{\hat{v}_{\alpha,i}\geq t,v_{\alpha_i}<t\}+1\{\hat{v}_{\alpha,i}<t,v_{\alpha,i}\geq t\}\Big]\\
&=\sup_{t\in[0,1]}\frac{1}{n}\sum_i\Big[1\{\hat{v}_{\alpha,i}\geq t,t-e<v_{\alpha_i}<t\}+1\{\hat{v}_{\alpha,i}<t,t\leq v_{\alpha,i}<t+e\}\Big]\\
&+\frac{1}{n}\sum_i\Big[1\{\hat{v}_{\alpha,i}\geq t,v_{\alpha,i}<t-e\}+1\{\hat{v}_{\alpha,i}<t,v_{\alpha,i}\geq t+e\}\Big]\\
&\leq\sup_{t\in[0,1]}\frac{1}{n}\sum_i 1\{t-e\leq v_{\alpha,i}\leq t+e\}+\frac{1}{ne}\sum_i|\hat{v}_{\alpha,i}-v_{\alpha,i}|\\
&\leq\sup_{t\in[0,1]}|H_0(t+e)-H_0(t-e)|+2\sup_{t\in[0,1]}|H_{n,0}(t)-H_0(t)|+\frac{1}{ne}\sum_i|\hat{v}_{\alpha,i}-v_{\alpha,i}|
\end{aligned}$$

for some $e>0$ arbitrarily small and bounded away from zero, the right hand side converges to zero in probability by results in Lemma A.1 and the uniform continuity of H_0. Using similar arguments in the proof of Theorem 6.1, we can establish that $\sup_{t\leq 1-\gamma}\left|\hat{Q}_n(t)-Q(t)\right|\xrightarrow{p}0$ and $\sup_{t\leq 1-\gamma}\left|\frac{\hat{V}_n(t)}{\hat{H}_{n,0}(t)}-Q(t)\right|\xrightarrow{p}0$ with $\hat{Q}_n(t)=\hat{H}_{n,1}(t)/\hat{H}_{n,0}(t)$ and $\hat{V}_n(t)=\frac{1}{n}\sum_{i=1}^n 1\{\hat{v}_{\alpha,i}\geq t\}1\{\theta_i\leq\theta_\alpha\}$ and consequently $\limsup_{n\to\infty}\mathbb{E}\Big[\frac{\hat{V}_n(\hat{\lambda}_n)}{\hat{H}_{n,0}(\hat{\lambda}_n)\bigvee 1}\Big]\leq\gamma$. ■

Proof. [Proof of Theorem 6.3]

We first show that $\hat{\lambda}_{1n}\xrightarrow{p}\lambda_1^*$ and $\hat{\lambda}_{2n}\xrightarrow{p}\lambda_2^*$, then by the continuous mapping theorem, we have $\hat{\lambda}_n=\max\{\hat{\lambda}_{1n},\hat{\lambda}_{2n}\}\xrightarrow{p}\max\{\lambda_1^*,\lambda_2^*\}=\lambda^*$. The second statement follows from Lemma A.3. The first statement holds because by the argument for Theorem 6.2, we have

$$\sup_{t\geq 1-\gamma}\left|\hat{Q}_n(t)-Q(t)\right|\xrightarrow{p}0 \tag{A.10}$$

And therefore for any $\epsilon>0$ not too large, we have $\inf_{t\leq\lambda^*-\epsilon}Q(t)>\gamma$ and $Q(\lambda^*+\epsilon)>\gamma$ by monotonicity of $Q(t)$. Combined with (A.10), we have $\hat{\lambda}_{1n}\xrightarrow{p}\lambda_1^*$.

Now define

$$
\begin{aligned}
H_{n,2} &= \frac{1}{n}\sum_{i=1}^{n} v_{\alpha,i}\mathbb{1}\{v_{\alpha,i} \geq t\} \\
\hat{H}_{n,2}(t) &= \frac{1}{n}\sum_{i=1}^{n} \hat{v}_{\alpha,i}\mathbb{1}\{\hat{v}_{\alpha,i} \geq t\} \\
\hat{U}_n(t) &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{\theta_i \geq \theta_\alpha, \hat{v}_{\alpha,i} \geq t\} \\
U_n(t) &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{\theta_i \geq \theta_\alpha, v_{\alpha,i} \geq t\} \\
H_2(t) &= \mathbb{P}(\theta_i \geq \theta_\alpha, v_{\alpha,i} \geq t) = \alpha\beta(t)
\end{aligned}
$$

It suffices to prove that $\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{\theta_i \geq \theta_\alpha, \hat{v}_{\alpha,i} \geq \hat{\lambda}_n\} \xrightarrow{p} H_2(\lambda^*)$. Using a similar argument as for Theorem 6.2, we can show

$$
\begin{aligned}
\sup_{t\in[0,1]} \left|\hat{H}_{n,2}(t) - H_2(t)\right| &\xrightarrow{p} 0 \\
\sup_{t\in[0,1]} \left|\hat{U}_n(t) - H_2(t)\right| &\xrightarrow{p} 0
\end{aligned}
$$

Combining this with the result that $\hat{\lambda}_n \xrightarrow{p} \lambda^*$, we have $\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{\theta_i \geq \theta_\alpha, \hat{v}_{\alpha,i} \geq \hat{\lambda}_n\} \xrightarrow{p} H_2(\lambda^*)$ by continuity of H_2. ■

Appendix B. A Discrete Bivariate Example

In this appendix we consider a case where G is a discrete distribution joint distribution in the pairs, θ, σ), in particular, $G(\theta, \sigma) = 0.85\delta_{(-1,6)} + 0.1\delta_{(4,2)} + 0.05\delta_{(5,4)}$. In contrast to the discrete example in Section 4.4, the unobserved variance σ^2 is now clearly informative about θ for this distribution G.

We will focus on the capacity constraint $\alpha = 0.05$ so $\theta_\alpha = 5$. For $T = 9$, the level curves for tail probability and posterior mean are shown in Figure B.1 and the selection set comparison for one sample realization in Figure B.2. The right panel of the figure plots the selection boundaries for the two ranking criteria for $\gamma = 10\%$. The non-monotonicity of $v_\alpha(y, s)$ in both y and s is apparent. The posterior mean criteria based on $\mathbb{E}(\theta|Y, S)$, prefers individuals with smaller variances compared to the rule based on the tail probability. Since sample variances S are informative about θ, when the sample variance is small and selection is based on posterior tail probability, the oracle is aware that such a small sample variance is only likely when $\theta = 4$, hence will only make a selection when we observe a very large y. As a result, the Oracle sets a higher selection threshold on y to avoid selecting individuals with true effect $\theta = 4$. On the other hand, the posterior mean criterion also tries to use information from the sample variance, but not as effectively for our selection objective. This can be seen in the level curves in the middle panel of Figure B.1. When the sample variance is small, the posterior mean shrinks very aggressively towards 4, thereby sacrificing valuable information from y. For a wide range of values for the sample mean y, the posterior mean delivers a value close to 4, thus failing to distinguish between those with $\theta = 5$ and those with $\theta = 4$. Consequently, the posterior mean rule sets a lower thresholding value on y for the selection region when sample variance is small resulting in inferior power performance, as shown in Table B.1.

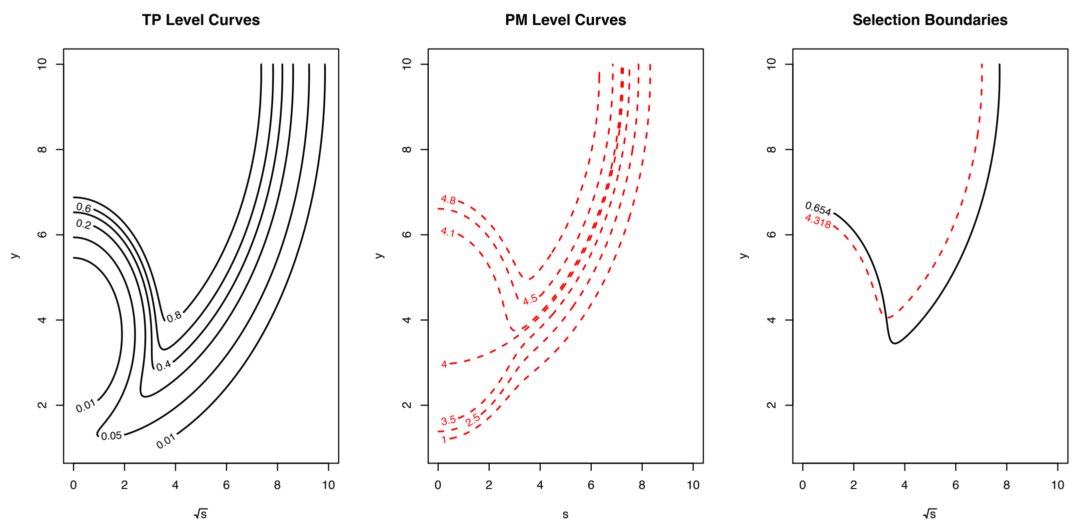

Figure B.1. The left panel plots the level curves for the posterior tail probability criterion and the middle panel depicts the level curves for posterior mean criterion. The right panel plots the selection boundary based on posterior mean ranking (shown as the red dashed lines) and the posterior tail probability ranking (shown as the black solid lines) for $\alpha = 5\%$ and $\gamma = 10\%$ with $G(\theta, \sigma^2)$ follows a three points discrete distribution.

Table B.1 reports several performance measures over 200 simulation repetitions with $n = 50,000$. There, we consider four additional methods for ranking:

- MLE: ranking of the maximum likelihood estimators, Y_i for each of the θ_i,

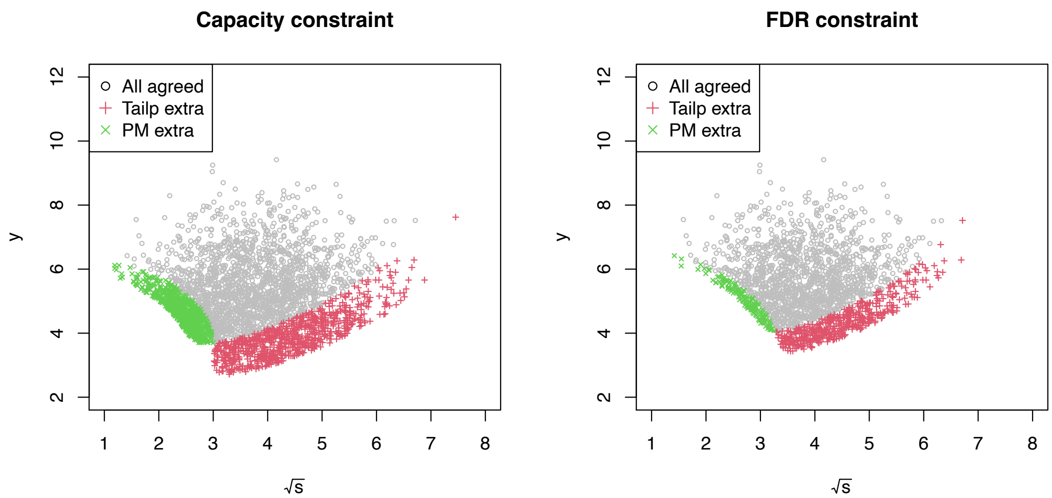

FIGURE B.2. Selection set comparison for one sample realization from the three points discrete distribution model: The left panel shows in black circles the agreed selected elements by both the posterior mean and posterior tail probability criteria under the capacity constraint, extra elements selected by the posterior mean are marked in green and extra elements selected by the posterior tail probability rule are marked in red. The right panel shows the comparison of the selected sets under both the capacity and FDR constraint with $\alpha = 5\%$ and $\gamma = 10\%$.

	$\gamma = 1\%$			$\gamma = 5\%$			$\gamma = 10\%$		
	Power	FDR	SelProp	Power	FDR	SelProp	Power	FDR	SelProp
PM	0.217	0.010	0.011	0.482	0.050	0.025	0.580	0.100	0.032
TP	0.252	0.010	0.013	0.561	0.050	0.030	0.697	0.100	0.039
P-value	0.651	0.349	0.050	0.651	0.350	0.050	0.651	0.349	0.050
MLE	0.611	0.390	0.050	0.611	0.390	0.050	0.610	0.390	0.050
PM-NIX	0.611	0.390	0.050	0.611	0.390	0.050	0.610	0.390	0.050
TP-NIX	0.619	0.382	0.050	0.619	0.382	0.050	0.618	0.382	0.050

TABLE B.1. Performance comparison for ranking procedures based on posterior mean, posterior tail probability, the P-value and the MLE of θ_i. All results are based on 200 simulation repetitions with n = 50,000 for G following the three point discrete distribution and T = 9 or when G is assumed to follow a normal-inverse-chi-square distribution. For the first two rows, number reported in the table correspond to performance when both capacity and FDR constraints are in place. For the last four rows, only capacity constraint is in place.

- P-values: ranking of the P-values of the conventional one-sided test of the null hypothesis $H_0 : \theta < \theta_\alpha$,
- PM-NIX: ranking of the posterior means based on the normal-inverse-chi-square (NIX) prior distribution,
- TP-NIX: ranking of the posterior tail probability based on NIX prior distribution

The first two of these selection rules ignore the compound decision perspective of the problem entirely. The other two ranking criteria we consider are based on posterior mean and tail probability assuming G follows a normal-inverse-chi-square (NIX) distribution (denoted as PM-NIX and TP-NIX in Table B.1). The parameters of the NIX distribution are estimated from the data, hence these rules can be viewed as generalization of James-Stein estimator for homogeneous variances and the Efron-Morris shrinkage estimator for known heterogeneous variances case. We refer the details of the NIX distribution and the posterior distribution of (θ, σ^2) to Example 5.1.

We report the power and false discovery rates for the posterior mean (denoted as PM) and posterior tail probability (denoted as TP) selection as well as the proportion of selected observations for $\alpha = 5\%$ and for several different γ under both capacity and FDR control. For all other four selection rules we only impose the capacity constraint, as how they are usually implemented in current practise. For the PM and TP rules, from the proportion selected observations we can infer whether the FDR constraint or the capacity constraint is binding in each configuration. Ranking based on the posterior tail probability clearly has better power performance for each of configurations when compared to the posterior mean ranking. When selecting as few as 5%, FDR constraints are binding for both PM and TP rule for all ranges of $\gamma \in \{1\%, 5\%, 10\%\}$. Among all the other rules, we see that the false discovery rate is around 40% and PM-NIX has identical performance as ranking based on the MLE for θ; this can be understood by noting that the posterior mean of θ under the NIX prior is simply linear shrinkage of the MLE of θ, hence it does not alter individual rankings between the two methods. TP-NIX behaves similarly to PM-NIX, with slightly better power and slightly lower false discovery rate.

APPENDIX C. A COUNTEREXAMPLE: NON-NESTED SELECTION REGIONS

Thus far we have stressed conditions under which selection regions are nested with respect to α, that is, for $\alpha_1 < \alpha_2 < \cdots < \alpha_m$, we have $\Omega_{\alpha_1} \subseteq \Omega_{\alpha_2} \subseteq \cdots \subseteq \Omega_{\alpha_m}$, for the selection regions. However, this need not hold when there is variance heterogeneity and when nesting fails we can have seemingly anomalous situations in which units are selected by the tail probability rule at some stringent, low α, but are then rejected for some less stringent, larger α's. To illustrate this phenomenon we will neglect the FDR constraint and focus on our discrete mixing distribution, $G = 0.85\delta_{-1} + 0.10\delta_2 + 0.05\delta_5$, with $\sigma \sim U[1/2, 4]$. The selection regions are depicted in Figure C.1 for $\alpha \in \{0.04, 0.05, 0.06, 0.08\}$. Units are selected when their observed pair, (y_i, σ_i) lies above these curves for various α's. When σ is small we see, as expected, that selection is nested: if a unit is selected at low α it stays selected at larger α's. However, when $\sigma = 3$, we see that there are units selected at $\alpha = 0.05$ and even $\alpha = 0.04$ and yet they are rejected for $\alpha = 0.06$. How can this be?

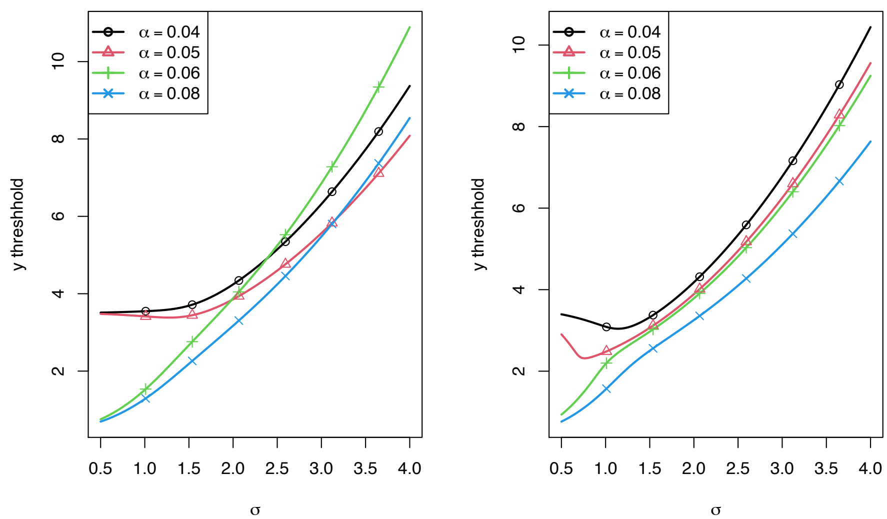

FIGURE C.1. Oracle selection boundaries (with just capacity constraint) for several α levels for the tail probability criterion (left panel) and posterior mean criterion (right panel) with a discrete example with $G = 0.85\delta_{-1} + 0.10\delta_2 + 0.05\delta_5$, and $\sigma \sim U[1/2, 4]$. Crossing of the boundaries implies that the selection regions are non-nested as explained in the text.

Imagine you are the oracle, so you know G, and when you decide to select with $\alpha = 0.06$ you know that you will have to select a few $\theta = 2$ types, since there are only 5 percent of the $\theta = 5$ types. Your main worry at that point is to try to avoid selecting any $\theta = -1$ types; this can be accomplished but only by avoiding the high σ types. In contrast when $\alpha = 0.05$ so we are trying to vacuum up all of the $\theta = 5$ types it is worth taking more of a risk with high σ types as long as their y_i is reasonably large.

The crossing of the selection boundaries and non-nestedness of the selection regions is closely tied up with the tail probability criterion and the α dependent feature of the hypothesis. If we repeat our exercise with the same G, and σ distribution, but select according to posterior means, we get the nested selection boundaries illustrated in Figure C.1. Proposition C.1 establishes this to be a general phenomenon for any distributution G.

It should be noted that the crossing of selection boundaries we have illustrated seems to have been anticipated in Henderson and Newton (2016), who consider similar tail criteria. They propose a ranking scheme that assigns rank equal to the smallest α for which a unit would be selected as a way to resolve the ambiguities generated by crossing. We don't see a compelling decision theoretic rationale for this revised ranking rule, instead we prefer to maintain some separation between the ranking and selection problems and focus on risk assessment as a way to reconcile them.

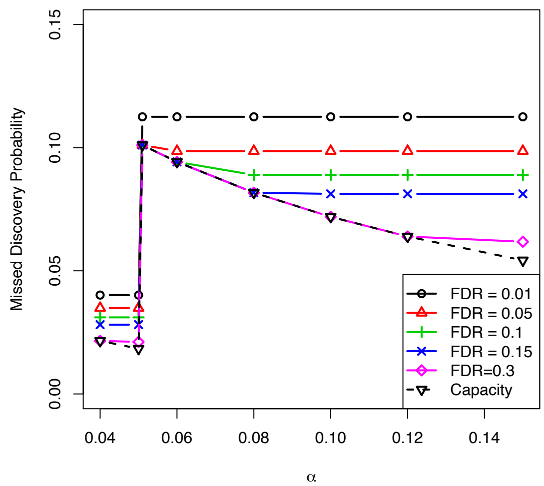

Figure C.2. Oracle risk evaluation for several α levels for the tail probability criterion and a discrete example with $G = 0.85\delta_{-1} + 0.10\delta_2 + 0.05\delta_5$, and $\sigma \sim U[1/2, 4]$. The solid lines correspond to the evaluation of the loss function specified in (3.1) with both capacity and FDR control constraints. The dotted line corresponds to oracle risk when only the capacity constraint is imposed.

The risk based on the loss function defined in (3.1) clearly depends on α and γ. More specifically it consists of three pieces, the leading term has the interpretation of "missed discovery" probability, which we try to minimize and the second and third pieces correspond to the FDR and capacity constraints respectively each weighted by a Lagrangian multiplier. Focusing on the first term, we have

$$\begin{aligned}\mathbb{E}[H_i(1-\delta_i)] &= \mathbb{E}[\mathbb{1}\{\theta_i \geq \theta_\alpha\}(1-\delta_i)] \\ &= \mathbb{P}[\theta_i \geq \theta_\alpha] - \mathbb{E}[\delta_i \mathbb{1}\{\theta_i \geq \theta_\alpha\}] \\ &= \mathbb{P}[\theta_i \geq \theta_\alpha] - \int \int_{\theta_\alpha}^{+\infty} \Big[1 - \Phi((t_\alpha(\lambda,\sigma) - \theta)/\sigma)\Big] dG(\theta) dH(\sigma)\end{aligned}$$

where λ, depends on α and γ, is determined by either the false discovery rate control or the capacity constraint, whichever binds.

A feature of discrete mixing distributions, G, is that the first term in the loss, $\mathbb{P}(\theta_i \geq \theta_\alpha)$, is piece-wise constant, with jumps occurring only at discontinuity points of G, while the second term depends on both θ_α and the cut-off values $t_\alpha(\lambda, \sigma)$. When the capacity constraint binds there exist ranges of α such that θ_α remains constant, while $t_\alpha(\lambda, \sigma)$ decreases for each σ, hence the risk with just capacity constraint binding is a decreasing function for α in the interval $(0.05, 0.15)$. On the other hand, when the FDR constraint binds, it can be shown that the cut-off value $t_\alpha(\lambda, \sigma)$ is constant. To see this recall that the cutoff λ determined by the FDR constraint is defined as $\mathbb{E}[(1 - v_\alpha(Y, \sigma))\mathbb{1}\{v_\alpha(Y, \sigma) \geq \lambda\}] = \gamma\mathbb{P}(v_\alpha(Y, \sigma) \geq \lambda)$, so when θ_α is constant over a range of α the distribution of $v_\alpha(Y, \sigma)$ does not change, and consequently the value λ is constant over that range of α.

Figure C.2 evaluates risk based on the optimal selection rule for various α and FDR levels, $\gamma \in \{0.01, 0.05, 0.1, 0.15, 0.3\}$. The solid curves correspond to risk evaluated at the optimal Bayes rule defined in Proposition 4.1. The dotted line corresponds to the risk evaluated at the Bayes rule when only the capacity constraint is imposed. As γ increases, the risk decreases as expected. For FDR levels as stringent as $\gamma = 0.01$, the FDR constraint binds and the risk is piece-wise constant. As the FDR level is relaxed, there are range of α such that capacity constraint becomes binding, and the risk decreases after the initial jump at $\alpha = 0.05$.

The evaluation of the risk for this particular example indicates that it is easier to select the top 5% individuals, those with $\theta = 5$. As we intend to select more in the right tail, we are facing more uncertainty. This also motivates a more systematic choice of (α, γ). Although selection based on the tail probability criterion can lead to non-nested selection regions, we conclude this sub-section by demonstrating that posterior mean selection is necessarily nested.

Proposition C.1. *Let the density function of y conditional on θ and σ be denoted as $f(y|\theta, \sigma)$. If selection is based on the posterior mean, $\delta_i = \{M(y, \sigma) \geq c(\alpha, \gamma)\}$ with*

$$M(y, \sigma) := \frac{\int \theta f(y|\theta, \sigma) dG(\theta)}{\int f(y|\theta, \sigma) dG(\theta)}$$

and $c(\alpha, \gamma)$ is chosen to satisfy both the capacity constraint at level α and the FDR constraint at level γ, then the selection regions, defined as $\Lambda_{\alpha,\gamma} = \{(y, \sigma) : M(y, \sigma) \geq c(\alpha, \gamma)\}$, are nested, that is, for any $\alpha_1 > \alpha_2$, $\Lambda_{\alpha_2,\gamma} \subseteq \Lambda_{\alpha_1,\gamma}$.