The original constrained problem

The original constrained problem is stated as

$$\min_{\tau} \iint -\text{sgn}\,(v)\, u\, f_Y\,(u, v; \tau)\; du\, dv$$
$$\text{subject to } \kappa - \iint \text{sgn}\,(v)\, w\, f_Z\,(w, v; \tau)\; dw\, dv \geq 0, \text{and } \tau^\intercal \tau = 1.$$

Suppose that the strict feasible set strict($\mathcal{F}$) is non-empty, and let τ^0 denote a constrained minimizer of this original problem. For simplicity, we let $g(\tau) = -\iint \text{sgn}(v)u\, f_Y(u, v; \tau)\, du\, dv$, and let $c(\tau) = \kappa - \iint \text{sgn}(v)w f_Z(w, v; \tau)\, dw\, dv$. The value of $g(\tau)$ at $\tau = \tau^0$, $g(\tau^0)$, is denoted by g^0. Similarly, c^0 denotes the value of $c(\tau)$ at $\tau = \tau^0$, $c(\tau^0)$. Also, $\mathcal{A}^0$ denotes the set of active constraint at τ^0, $\mathcal{A}(\tau^0)$. In our current case, it is either $\mathcal{A}^0 = \{c^0\}$, or $\mathcal{A}^0 = \varnothing$.

Log-barrier penalty function

One of the method to solve constrained optimization problem is use log-barrier penalty function, which is a composite measure of the objective function and the penalty of violating the constraints. The log-barrier penalty function is then formalized as

$$B(\tau, \mu) = \iint -\text{sgn}(v)u f_Y(u, v; \tau)\, du\, dv - \mu \ln \left[\kappa - \iint \text{sgn}(v) w f_Z(w, v; \tau)\, dw\, dv\right],$$

where μ is a sequence of decreasing positive very small constants converging to zero. Let τ^* denote an unconstrained minimizer of f $B(\tau, \mu)$ as $\tau^*(\mu)$ for emphasizing that it is a vector function of μ, or τ^*_μ for short. It can be proven that the constraint is strictly satisfied, i.e., $c(\tau^*_\mu) = \kappa - \iint \text{sgn}\,(v)\, w\, f_Z\left(w, v; \tau^*_\mu\right)\; dw\, dv > 0$. The gradient of $B(\tau, \mu)$ is

$$\nabla B(\tau, \mu) = \iint -\text{sgn}(v)u\nabla f_Y(u, v; \tau)\, du\, dv + \mu \frac{\iint \text{sgn}(v) w \nabla f_Z(w, v; \tau)\, dw\, dv}{\kappa - \iint \text{sgn}(v) w f_Z(w, v; \tau)\, dw\, dv},$$

noting that ∇ represents the first order derivative with respect to τ.

If we are willing to assume that $\nabla B(\tau, \mu)$ is twice-continuously differentiable, it must hold that $\nabla B(\tau_\mu, \mu) = 0$, which means that

$$\iint \text{sgn}\,(v)\, u\, \nabla f_Y\,(u, v; \tau_\mu)\; du\, dv = \mu \frac{\iint \text{sgn}(v) w \nabla f_Z(w, v; \tau)\, dw\, dv}{\kappa - \iint \text{sgn}(v) w f_Z(w, v; \tau_\mu)\, dw\, dv}$$

The barrier multiplier, the coefficient in this linear relationship above, denoted by λ_μ, is defined as

$$\lambda_\mu \triangleq \frac{\mu}{\kappa - \iint \text{sgn}(v) w f_Z(w, v; \tau_\mu)\, dw\, dv}.$$

This relationship can be re-written as

$$\lambda_\mu \left[\kappa - \iint \operatorname{sgn}(v) w f_Z(w, v; \tau_\mu)\, dw\, dv \right] = \mu.$$

This relationship between the barrier multiplier, the constraint value, and the barrier parameter, called perturbed complementarity, is analogous as $\mu \to 0$ to the complementarity condition $c(\tau^*)\lambda^* = 0$ that holds at a KKT point.

Optimality conditions for the central path/barrier trajectory

Before moving forward to estimate the log-barrier function, we need to examine the optimality conditions for the central path/barrier trajectory, which ensure that $\lim_{\mu \to 0_+} \tau^*(\mu) = \tau^0$. Consider the problem stated above. Let $\mathcal{F}$ denote the feasible region, and assume that the set strict$(\mathcal{F})$ of strictly feasible points is non-empty. Let τ^0 be a local constrained minimizer, with g^0 denoting $g(\tau^0) = \nabla f(\tau^0)$, J^0 denoting $J(\tau^0) = \nabla c(\tau^0)$, and so on, and let $\mathcal{A}$ denote $\mathcal{A}(\tau^0)$. Assume theat the following sufficient optimality conditions hold at τ^0:

(a) τ^0 is a KKT point, i.e., there exists a nonempty set $\mathcal{M}_\lambda$ of Lagrange multipliers λ satisfying
$$\mathcal{M}_\lambda = \{\lambda : g^0 = \lambda^T J^0, \lambda \geq 0, \text{and } c(\tau^0) \cdot \lambda = 0\}$$

(b) the MFCQ (a condition on the constraints) holds at τ^0, i.e., there exists p such that $J^0_{\mathcal{A}} p > 0$, where $J^0_{\mathcal{A}}$ denotes the Jacobian of the active constraints at τ^0; and

(c) there exists $\omega > 0$, such that $p^T H(\tau^0, \lambda) p \geq w \|p\|^2$ for all $\lambda \in \mathcal{M}_\lambda$ and all nonzero p satisfying $g^{0T} p = 0$ and $J^0_{\mathcal{A}} p \geq 0$, where $H(x^0, \lambda)$ is the Hessian of the Lagrangian (2.11). $H(\tau, \lambda) \triangleq \nabla^2_{\tau\tau} L(\tau, \lambda) = \nabla^2 f(\tau) - \sum_{i=1}^m \lambda_i \nabla^2 c_i(\tau)$

Assume that a logrithmic barrier method is applied in which μ_k converges monotonically to zero as $k \to \infty$. Then,

(i) there is at least one subsequence of unconstrained minimizers of the barrier function $B(\tau, \mu_k)$ converging to τ^0;

(ii) let $\{\tau^k\}$ denote such a convergent subsequence, with the obvious notation that c_i^k denotes $c_i(\tau^k)$, and so on. Then the sequence of barrier multipliers $\{\lambda^k\}$, whose i-th component is μ_k / c_i^k, is bounded;

(iii) $\lim_{k \to \infty} \lambda^k = \bar{\lambda} \in \mathcal{M}_\lambda$

If, in addition, strict complementarity holds at τ^0, i.e, there is a vector $\lambda \in \mathcal{M}_\lambda$ such that $\lambda_i > 0$ for all $i \in \mathcal{A}$, then

(iv) $\bar{\lambda}_{\mathcal{A}} > 0$;

(v) for sufficiently large k, the Hessian matrix $\nabla^2 B(\tau^k, \mu_k)$ is positive defnite;

(vi) a unique, continuously differentiable vector function $\tau(\mu)$ of unconstrained minimizers of $B(\tau, \mu)$ exisits for positive μ in a neighborhood of $\mu = 0$; and

(vii) $\lim_{\mu \to 0_+} \tau^*(\mu) = \tau^0$.

Suppose there is the subsequence $\{\mu_k\}$ corresponding to the convergent subsequence $\{\tau^k\}$. It can be proven that $c(\tau^k) > 0$. Assume all the conditions above hold, we can then estimate the penalty function and solve it using unconstrained optimization method to estimate the log-barrier minimizer trajectory, denoted by $\{\widehat{\tau}^k\}$.

Estimation of the log-barrier penalty function
To estimate the log-barrier penalty function, we use kernel density estimators, denoted by $\widehat{f}_Y(u, v; \tau)$ and $\widehat{f}_Z(w, v; \tau)$, to estimate the corresponding density functions. Hence, the estimated log-barrier function is

$$\widehat{B}(\tau, \mu) = \iint -\mathrm{sgn}(v) u \widehat{f}_Y(u, v; \tau)\, du\, dv - \mu \ln \left[\kappa - \iint \mathrm{sgn}(v) w \widehat{f}_Z(w, v; \tau)\, dw\, dv\right],$$

and the gradient of the estimator is

$$\begin{aligned}\nabla \widehat{B}(\tau, \mu) &= \iint -\mathrm{sgn}(v) u \nabla \widehat{f}_Y(u, v; \tau)\, du\, dv + \mu \frac{\iint \mathrm{sgn}(v) w \nabla \widehat{f}_Z(w, v; \tau)\, dw\, dv}{\kappa - \iint \mathrm{sgn}(v) w \widehat{f}_Z(w, v; \tau)\, dw\, dv} \\ &= \iint -\mathrm{sgn}(v) u \nabla \widehat{f}_Y(u, v; \tau)\, du\, dv + \widehat{\lambda}_\mu \iint \mathrm{sgn}(v) w \nabla \widehat{f}_Z(w, v; \tau)\, dw\, dv,\end{aligned}$$

where $\widehat{\lambda}_\mu(\tau) = {}^{\mu}/_{\kappa - \iint \mathrm{sgn}(v) w \widehat{f}_Z(w, v; \tau)\, dw\, dv}$.

Consistency of $\widehat{\tau}^k$ and $\widehat{\lambda}_\mu$.
We need to prove that $\widehat{\tau}^k$ is a consistent estimator of τ^{*k}.
$\widehat{\tau}^k - \tau^{*k} = O_p(n^{1/2})$, and $\widehat{\lambda}^k - \lambda^{*k} = O_p(n^{1/2})$.
Theorem proved that λ_μ is bounded.

Asymptotic distribution of $\widehat{\tau}^k$
Estimating equations:

$$\nabla \widehat{B}(\tau, \mu) = \iint -\mathrm{sgn}(v) u \nabla \widehat{f}_Y(u, v; \tau)\, du\, dv + \hat{\lambda}_\mu(\tau) \iint \mathrm{sgn}(v) w \nabla \widehat{f}_Z(w, v; \tau)\, dw\, dv = 0$$

where $\hat{\lambda}_\mu(\tau) = {}^{\mu}/_{[\kappa - \iint \mathrm{sgn}(v) w \widehat{f}_Z(w, v; \tau)\, dw\, dv]}$.

$$
\begin{aligned}
\nabla\widehat{B}(\tau,\mu) =& \iint -\mathrm{sgn}(v)u\nabla\widehat{f}_Y(u,v;\tau)\,du\,dv + \hat{\lambda}_\mu \iint \mathrm{sgn}(v)w\nabla\widehat{f}_Z(w,v;\tau)\,dw\,dv \\
=& -\frac{2}{n}\sum_{i=1}^n \boldsymbol{X}_{i,1}^\intercal\boldsymbol{\beta}_{Y1}k\left(-\frac{\boldsymbol{X}_i^\intercal\boldsymbol{\tau}}{h}\right)\boldsymbol{X}_i + \hat{\lambda}_\mu(\tau)\frac{2}{n}\sum_{i=1}^n \boldsymbol{X}_{i,1}^\intercal\boldsymbol{\beta}_{Z1}k\left(-\frac{\boldsymbol{X}_i^\intercal\boldsymbol{\tau}}{h}\right)\boldsymbol{X}_i \\
=& N(\mu_1,\Sigma_1) + C_pN(\mu_2,\Sigma_2)
\end{aligned}
$$

$$
\begin{aligned}
\nabla^2\widehat{B}(\tau,\mu) =& \iint -\mathrm{sgn}(v)u\nabla^2\widehat{f}_Y(u,v;\tau)\,du\,dv + \\
& \nabla\hat{\lambda}_\mu(\tau)\iint \mathrm{sgn}(v)w\nabla\widehat{f}_Z(w,v;\tau)\,dw\,dv + \hat{\lambda}_\mu(\tau)\iint \mathrm{sgn}(v)w\nabla^2\widehat{f}_Z(w,v;\tau)\,dw\,dv \\
=& -\frac{2}{nh}\sum_{i=1}^n \boldsymbol{X}_{i,1}^\intercal\left(\hat{\lambda}_\mu(\tau)\boldsymbol{\beta}_{Z1} - \boldsymbol{\beta}_{Y1}\right)k'\left(-\frac{\boldsymbol{X}_i^\intercal\boldsymbol{\tau}}{h}\right)\boldsymbol{X}_i\boldsymbol{X}_i^\intercal + \\
& \frac{2}{n}\sum_{i=1}^n \nabla\hat{\lambda}_\mu(\tau)\boldsymbol{X}_{i,1}^\intercal\boldsymbol{\beta}_{Z1}k\left(-\frac{\boldsymbol{X}_i^\intercal\boldsymbol{\tau}}{h}\right)\boldsymbol{X}_i
\end{aligned}
$$

$$
\begin{aligned}
\nabla\widehat{\lambda}_\mu(\tau) =& \frac{\mu}{\left(\kappa - \iint \mathrm{sgn}(v)w\widehat{f}_Z(w,v;\tau)\,dw\,dv\right)^2}\iint \mathrm{sgn}(v)w\nabla\widehat{f}_Z(w,v;\tau)\,dw\,dv \\
=& \mu\left[\kappa - \frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_{i,1}^\intercal\boldsymbol{\beta}_{Z1}\left\{1 - 2K\left(-\frac{\boldsymbol{X}_i^\intercal\boldsymbol{\tau}}{h}\right)\right\}\right]^{-1}\left[\frac{2}{n}\sum_{i=1}^n \boldsymbol{X}_{i,1}^\intercal\boldsymbol{\beta}_{Z1}k\left(-\frac{\boldsymbol{X}_i^\intercal\boldsymbol{\tau}}{h}\right)\boldsymbol{X}_i\right]
\end{aligned}
$$

[Notation: k and κ looks to similar]
[Need to estimate $\hat{\beta}$ too]
Need to prove that the difference between $\widehat{B}_n(\tau,\hat{\beta},\mu)$ and $\widehat{B}_n(\tau,\beta^*,\mu)$ is negligible? i.e., $\widehat{B}_n(\beta^*) - \widehat{B}_n(\hat{\beta}) = O_p(n^{-1/2})$

Taylor expansion of $\nabla\widehat{B}(\tau^{*k},\mu)$ at $\tau = \hat{\tau}^k$ shows that

$$
\nabla\widehat{B}(\tau^{*k},\mu) = \nabla\widehat{B}(\hat{\tau}^k,\mu) - \nabla^2\widehat{B}(\tilde{\tau}^k,\mu)(\widehat{\tau}^k - \tau^{*k}),
$$

where $\tilde{\tau}^k$ is between τ^{*k} and $\widehat{\tau}^k$. As $\hat{\tau}^k$ is the minimizer of $B(\tau,\mu)$, it satisfies the first order condition that $\nabla B(\hat{\tau}^k,\mu) = 0$. Therefore, we have

$$
\sqrt{n}\nabla\widehat{B}(\tau^{*k},\mu) = -\sqrt{n}\nabla^2\widehat{B}(\tilde{\tau}^k,\mu)(\widehat{\tau}^k - \tau^{*k}).
$$

Derivation of the integrations
The integration we need

$$
\begin{aligned}
&\iint \operatorname{sgn}(v)\, u\, f(u, v; \tau, \beta_{\cdot 1})\, du\, dv \\
=&2 \iint u\, \mathbb{I}(v \geq 0)\, f(v, u; \tau, \beta_{\cdot 1})\, dv\, du - \int u\, f(u; \beta_{\cdot 1})\, du
\end{aligned}
$$

The estimator is

$$
\begin{aligned}
&\iint \operatorname{sgn}(v)\, u\, \widehat{f}_n(u, v; \tau, \beta_{\cdot 1})\, du\, dv \\
=&2 \iint u\, \mathbb{I}(v \geq 0)\, \widehat{f}_n(v, u; \tau, \beta_{\cdot 1})\, dv\, du - \int u\, \widehat{f}_n(u; \tau, \beta_{\cdot 1})\, du \\
=&\frac{2}{nh^2} \iint u\, \mathbb{I}(v \geq 0) \sum_{i=1}^{n} k\left(\frac{v - V_i}{h}\right) k\left(\frac{u - U_i}{h}\right) du\, dv - \\
&\frac{1}{nh} \int u \sum_{i=1}^{n} k\left(\frac{u - U_i}{h}\right) du \\
=&\frac{2}{n} \sum_{i=1}^{n} \boldsymbol{X}_{i,1}^{\intercal} \boldsymbol{\beta}_{\cdot 1} \left\{1 - K\left(-\frac{\boldsymbol{X}_i^{\intercal} \boldsymbol{\tau}}{h}\right)\right\} - \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_{i,1}^{\intercal} \boldsymbol{\beta}_{\cdot 1} \\
=&\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_{i,1}^{\intercal} \boldsymbol{\beta}_{\cdot 1} \left\{1 - 2K\left(-\frac{\boldsymbol{X}_i^{\intercal} \boldsymbol{\tau}}{h}\right)\right\}
\end{aligned}
$$

where $\widehat{f}_n(u_1, u_2; \tau, \widehat{\beta}_{\cdot 1})$ are the kernel density estimator for $(X^{\intercal}\tau, X^{\intercal}\beta_{\cdot 1})$ with the forms of

$$
\widehat{f}_n(u, v; \boldsymbol{\tau}, \widehat{\boldsymbol{\beta}}_{\cdot 1}) = \frac{1}{nh^2} \sum_{i=1}^{n} k\left(\frac{u - U_i}{h}\right) k\left(\frac{v - V_i}{h}\right).
$$

Moreover, $K(s)$ is the corresponding CDF of the kernel function $k(s)$, which is chosen to be a symmetric probability density. More precisely, $k(s)$ satisfies the following assumptions:

1. $\int_{-\infty}^{\infty} k(s)\, ds = 1$.
2. $k(s) > 0$ for all s.
3. $k(-s) = k(s)$ for all s.
4. The first order derivative of the kernel, $k'(s)$, exists and is bounded.

The last equality above holds by following the derivation.

We first derive $\frac{2}{h^2}\iint u_2\,\mathbb{I}\left(u_1\geq 0\right)k\left(\frac{u_1-U_{i,1}}{h}\right)k\left(\frac{u_2-U_{i,2}}{h}\right)\,du_1\,du_2$. Let $s=\frac{u_1-U_{i,1}}{h}$ and $t=\frac{u_2-U_{i,2}}{h}$. Then, $u_1=U_{i,1}+sh$ and $u_2=U_{i,2}+th$. Also, $du_1=h\,ds$ and $du_2=h\,dt$.

$$
\begin{aligned}
&\frac{2}{h^2}\iint u_2\,\mathbb{I}\left(u_1\geq 0\right)k\left(\frac{u_1-U_{i,1}}{h}\right)k\left(\frac{u_2-U_{i,2}}{h}\right)\,du_1\,du_2\\
=&2\iint\left(U_{i,2}+th\right)\mathbb{I}\left(U_{i,1}+sh\geq 0\right)k\left(s\right)k\left(t\right)\,ds\,dt\\
=&2\int U_{i,2}\,\mathbb{I}\left(s\geq-\frac{U_{i,1}}{h}\right)k\left(s\right)\,ds\\
=&2U_{i,2}\left\{1-K\left(-\frac{U_{i,1}}{h}\right)\right\}\\
=&2\boldsymbol{X}_{i,1}^{\intercal}\boldsymbol{\beta}_{\cdot 1}\left\{1-K\left(-\frac{\boldsymbol{X}_i^{\intercal}\boldsymbol{\tau}}{h}\right)\right\},
\end{aligned}
$$

where $K\left(s\right)=\int k\left(s\right)\,ds+c$. The second equality holds, as $\int k(t)\,dt=1$ and $\int t\,k(t)\,dt=0$. The third equality holds as $\int\mathbb{I}\left(s\geq-\frac{U_{i,1}}{h}\right)k\left(s\right)\,ds=1-\int_{-\infty}^{-U_{i,1}/h}k\left(s\right)\,ds=1-K\left(-\frac{U_{i,1}}{h}\right)$, where $U_{i,1}=\boldsymbol{X}_i^{\intercal}\boldsymbol{\tau}$.

Then, we derive $\frac{1}{h}\int u_2k(\frac{u_2-U_{i,2}}{h})\,du_2$ by changing variable similarly. Let $t=\frac{u_2-U_{i,2}}{h}$,and we get $u_2=U_{i,2}+th$, and $du_2=h\,dt$.

$$
\begin{aligned}
&\frac{1}{h}\int u_2k\left(\frac{u_2-U_{i,2}}{h}\right)\,du_2\\
=&\int\left(U_{i,2}+th\right)k\left(t\right)\,dt\\
=&U_{i,2}=\boldsymbol{X}_{i,1}^{\intercal}\boldsymbol{\beta}_{\cdot 1}.
\end{aligned}
$$

Again, the second equality holds as $\int k(t)\,dt=1$, and $\int t\,k(t)\,dt=0$.
The integration over the first-order derivative

$$
\begin{aligned}
&\iint\operatorname{sgn}\left(v\right)u\,\nabla\widehat{f}_n\left(u,v;\tau,\beta_{\cdot 1}\right)\,du\,dv\\
=&\frac{\partial}{\partial\tau}\iint\operatorname{sgn}\left(v\right)u\,\widehat{f}_n\left(u,v;\tau,\beta_{\cdot 1}\right)\,du\,dv\\
=&\frac{2}{n}\sum_{i=1}^{n}\boldsymbol{X}_{i,1}^{\intercal}\boldsymbol{\beta}_{\cdot 1}k\left(-\frac{\boldsymbol{X}_i^{\intercal}\boldsymbol{\tau}}{h}\right)\boldsymbol{X}_i
\end{aligned}
$$