Simpson's $\Rightarrow$ Robinson's Paradox

Based on inequalities:

Within-Pooled-Between inequality

$$\hat\beta_W < \hat\beta_P < \hat\beta_B$$
or
$$\hat\beta_W > \hat\beta_P > \hat\beta_B$$
or
$$\hat\beta_W = \hat\beta_P = \hat\beta_B$$

Simpson's
$\hat\beta_W \times \hat\beta_P < 0$

Robinson's
$\hat\beta_W \times \hat\beta_B < 0$

$\underset{n\times 1}{Y}$ $\underset{n\times p}{X}$ $\underset{n\times g}{G}$ matrix of Group indicators. $\underset{\sim}{1} \in \mathcal{L}(G)$

eventually $p = 1$

$$\begin{bmatrix} 1 & 0 & 0 & \cdots \\ 1 & 0 & 0 & \\ 1 & 0 & 0 & \\ 0 & 1 & 0 & \\ 0 & 1 & 0 & \\ & 1 & 0 & \\ & 1 & 0 & \\ \vdots & \vdots & \vdots & \end{bmatrix}$$

G includes intercept

Let $P_G = G(G'G)^{-1}G'$

$Q_G = I - P_G$

$\underset{n\times p}{\bar{X}_G} = \underset{n\times n}{P_G}\ \underset{n\times p}{X}$

$\underset{n\times p}{X - \bar{X}_G} = Q_G X$

Same as cvar(X, id)

Same as dvar(X, id)

CWG

Within: $Y = X\hat{\beta}_W + G\hat{\gamma} + e$ $\quad e \perp \mathcal{L}(X, G)$

or AVP $Q_G Y = Q_G X \hat{\beta}_W + e$

Between: $P_G Y = P_G X \hat{\beta}_B + d$ $\quad d \perp P_G X$

(weighted by n_i) $\quad X'd = 0$?

Pooled $Y = X\hat{\beta}_P + f$ $\quad f \perp X$

So

$$\left[\begin{array}{l} X'Y = X'X\hat{\beta}_P \\ X'Q_G Y = X'Q_G X\hat{\beta}_W \\ X'P_G Y = X'P_G X\hat{\beta}_B \end{array}\right.$$

① from Pooled
② from AVP
③ from Between

Why?

① since $\underset{\sim}{f} \perp \mathcal{L}(X)$ so $X'\underset{\sim}{f} = 0$

② since $\underset{\sim}{e} \perp \mathcal{L}(X, G)$ so $X'\underset{\sim}{e} = 0$

③ Since $d = P_G Y - P_G X\hat{\beta}_B$
it follows that $d \in \mathcal{L}(G)$.
By the projection theorem $\underset{\sim}{d} \perp \mathcal{L}(P_G X)$
So $0 = X'P_G d = X'd$

③ Since $\underset{\sim}{d} = P_G Y - P_G X\hat{\underset{\sim}{\beta}}_B$

it follows that $\underset{\sim}{d} \in \mathcal{L}(G)$

By the projection theorem

$$d \perp \mathcal{L}(P_G X)$$

So $X' P_G d = 0$

But $P_G \underset{\sim}{d} = \underset{\sim}{d}$ so $X' \underset{\sim}{d} = 0$

Now using LHS$_s$ ② + ③ = ① because $P_G + Q_G = I$

So using RHS$_s$

$$X'X\hat{\beta}_P = X' Q_G X \hat{\beta}_W + X' P_G X \hat{\beta}_B$$

$$(A+B)\hat{\beta}_P = A\hat{\beta}_W + B\hat{\beta}_B$$

$$\hat{\beta}_P = \frac{a\hat{\beta}_W + b\hat{\beta}_B}{a+b} \qquad a, b \geq 0$$

Noting that $X'X = X'Q_G X + X'P_G X$

$$\hat{\beta}_p = (A+B)^{-1}(A\hat{\beta}_W + B\hat{\beta}_B)$$

where
$A = X'Q_G X$
$B = X'P_G X$

So $\hat{\beta}_p$ is on the osculant path of the families of ellipses with shape

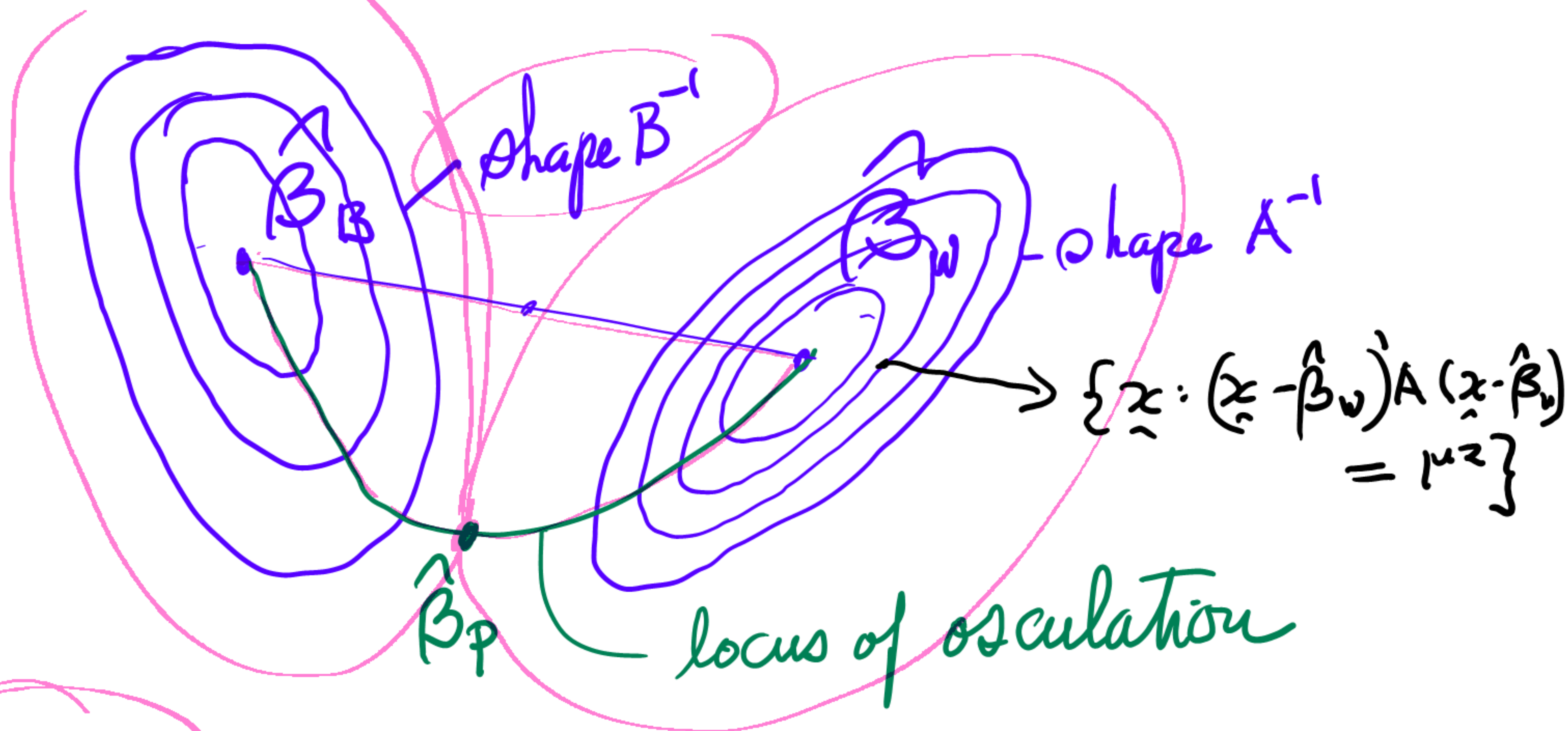
A^{-1} centered at $\hat{\beta}_W$ and

B^{-1} " " $\hat{\beta}_B$

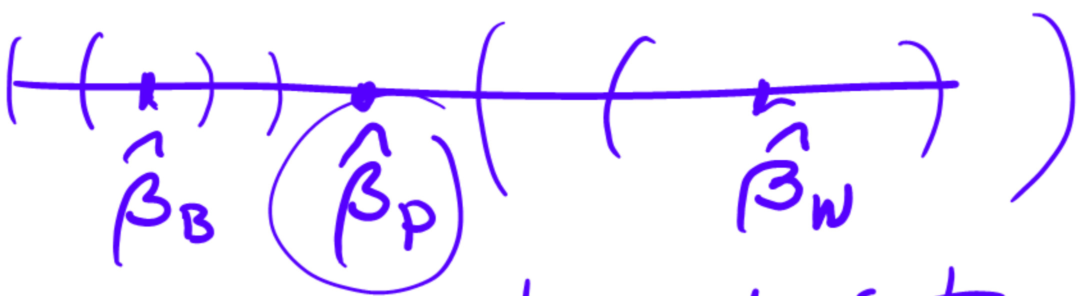
$\boxed{\hat{\beta}_p} = w_1 \boxed{\hat{\beta}_W} + w_2 \boxed{\hat{\beta}_B}$

$w_1 + w_2 = 1$

$w_i \geq 0$



If $p = 1$



So $\hat{\beta}_P$ is in convex hull of interval formed by $\hat{\beta}_B$ & $\hat{\beta}_W$

So $\hat{\beta}_W \le \hat{\beta}_p \le \hat{\beta}_B$

OR $\hat{\beta}_W \ge \hat{\beta}_p \ge \hat{\beta}_B$

Can $\hat{\beta}_P = \hat{\beta}_W \neq \hat{\beta}_B$?

If so $X' P_G X = 0 = X' P_G' P_G X$

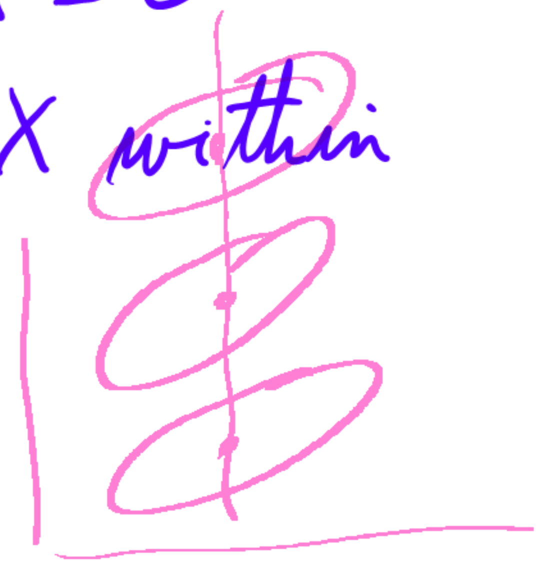
So $P_G X = 0$ and $\hat{\beta}_B$ is undefined because all X's have the same mean

β_B^{und}

Can $\hat{\beta}_P = \hat{\beta}_B \neq \hat{\beta}_W$?

Then $X' Q_G X = 0$ and $Q_G X = 0$

so there is no variability in X within groups and $\hat{\beta}_W$ is undefined.

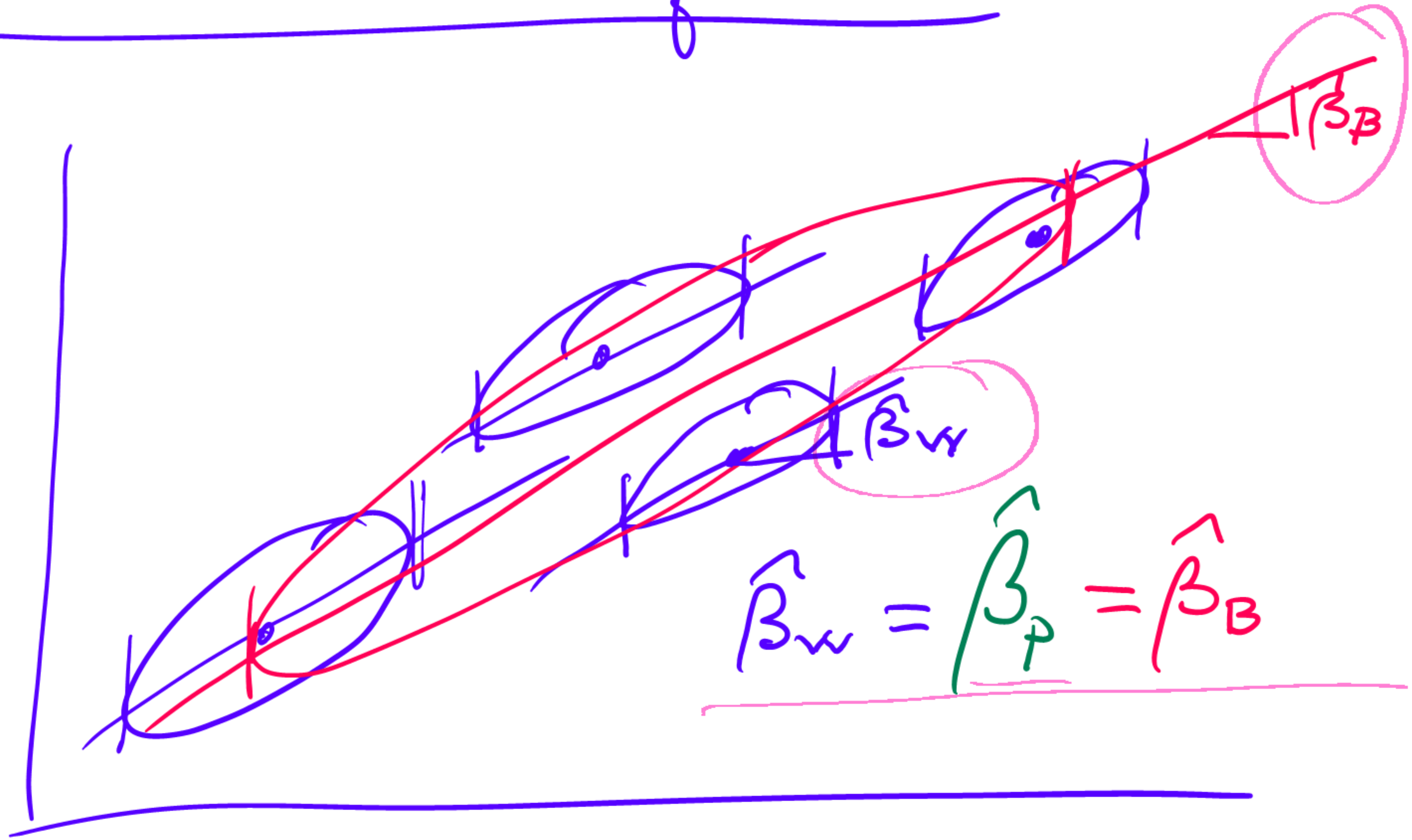


$\leq \quad \leq$

$> \quad \geq$

$= \quad =$

When are all 3 equal?



$$\hat{\beta}_W = \hat{\beta}_P = \hat{\beta}_B$$