高级计量经济学

Lecture 4: Frisch-Waugh-Lovell Theorem

黄嘉平

工学博士 经济学博士
深圳大学中国经济特区研究中心 讲师

办公室 粤海校区汇文楼1510
E-mail huangjp@szu.edu.cn
Website https://huangjp.com

分块回归

分块回归

Partitioned Regression

我们可以将解释变量分为两组，即

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 & \boldsymbol{X}_2 \end{bmatrix},$$

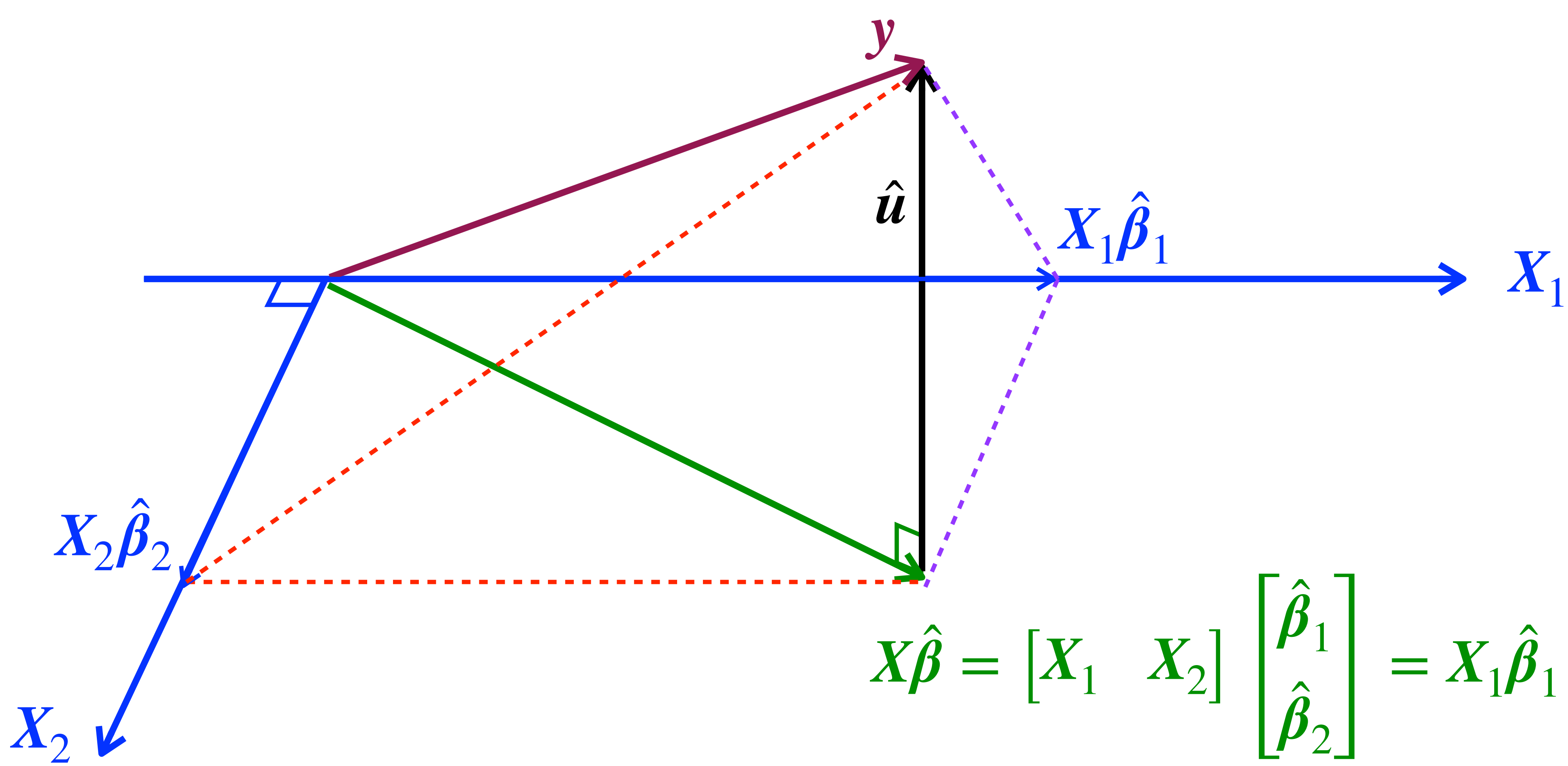
其中 $\boldsymbol{X}_1$ 为 $n \times k_1$，$\boldsymbol{X}_2$ 为 $n \times k_2$，$k_1 + k_2 = k$。回归模型可以写成

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{u}$$

接下来我们首先假设 $\boldsymbol{X}_1$ 中的所有变量与 $\boldsymbol{X}_2$ 中的所有变量正交，然后再放松这个假设。

若 $\boldsymbol{X}_1$ 中的变量与 $\boldsymbol{X}_2$ 中的变量正交，则 $\boldsymbol{X}_1^\top\boldsymbol{X}_2 = \boldsymbol{O}$。

当 $\boldsymbol{X}_1$ 与 $\boldsymbol{X}_2$ 正交



$$\boldsymbol{X}\hat{\boldsymbol{\beta}} = \begin{bmatrix}\boldsymbol{X}_1 & \boldsymbol{X}_2\end{bmatrix}\begin{bmatrix}\hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2\end{bmatrix} = \boldsymbol{X}_1\hat{\boldsymbol{\beta}}_1 + \boldsymbol{X}_2\hat{\boldsymbol{\beta}}_2$$

用 $\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{u}$ 拟合的 $\hat{\boldsymbol{\beta}}_1$ 和 $\hat{\boldsymbol{\beta}}_2$ 分别等于
用 $\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{u}_1$ 和 $\boldsymbol{y} = \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{u}_2$ 拟合的参数值。

如果 $\boldsymbol{X}_1$ 和 $\boldsymbol{X}_2$ 正交，在研究其中之一时可以不控制另一组变量。

正交向量与样本的 Pearson 相关系数

$\boldsymbol{x}^\top \boldsymbol{y} = 0$ 和 $r_{xy} = 0$ 是经常容易混淆的两个概念。

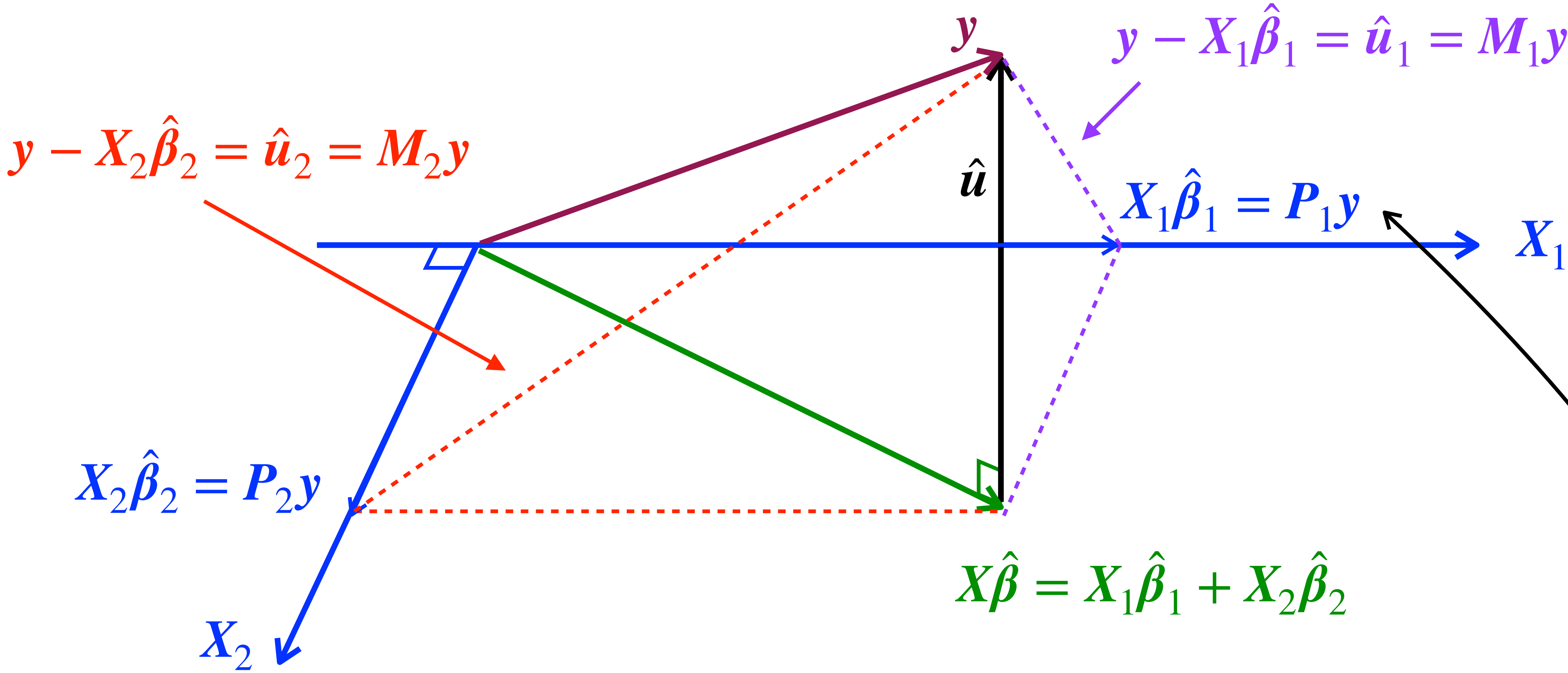
样本相关系数 r_{xy} 的定义是

$$r_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

可见，只有在 $\bar{x}\bar{y} = 0$ 时，$\boldsymbol{x}^\top \boldsymbol{y} = 0 \;\Leftrightarrow\; r_{xy} = 0$。

若将变量进行中心化处理，即 $x_i - \bar{x}$，则处理后的样本均值为零，此时**正交**等同于**不相关**。（相同结论在总体层面也成立）

$\boldsymbol{y}$ 在 $\mathcal{S}(\boldsymbol{X}_1)$ 和 $\mathcal{S}(\boldsymbol{X}_2)$ 上的投影



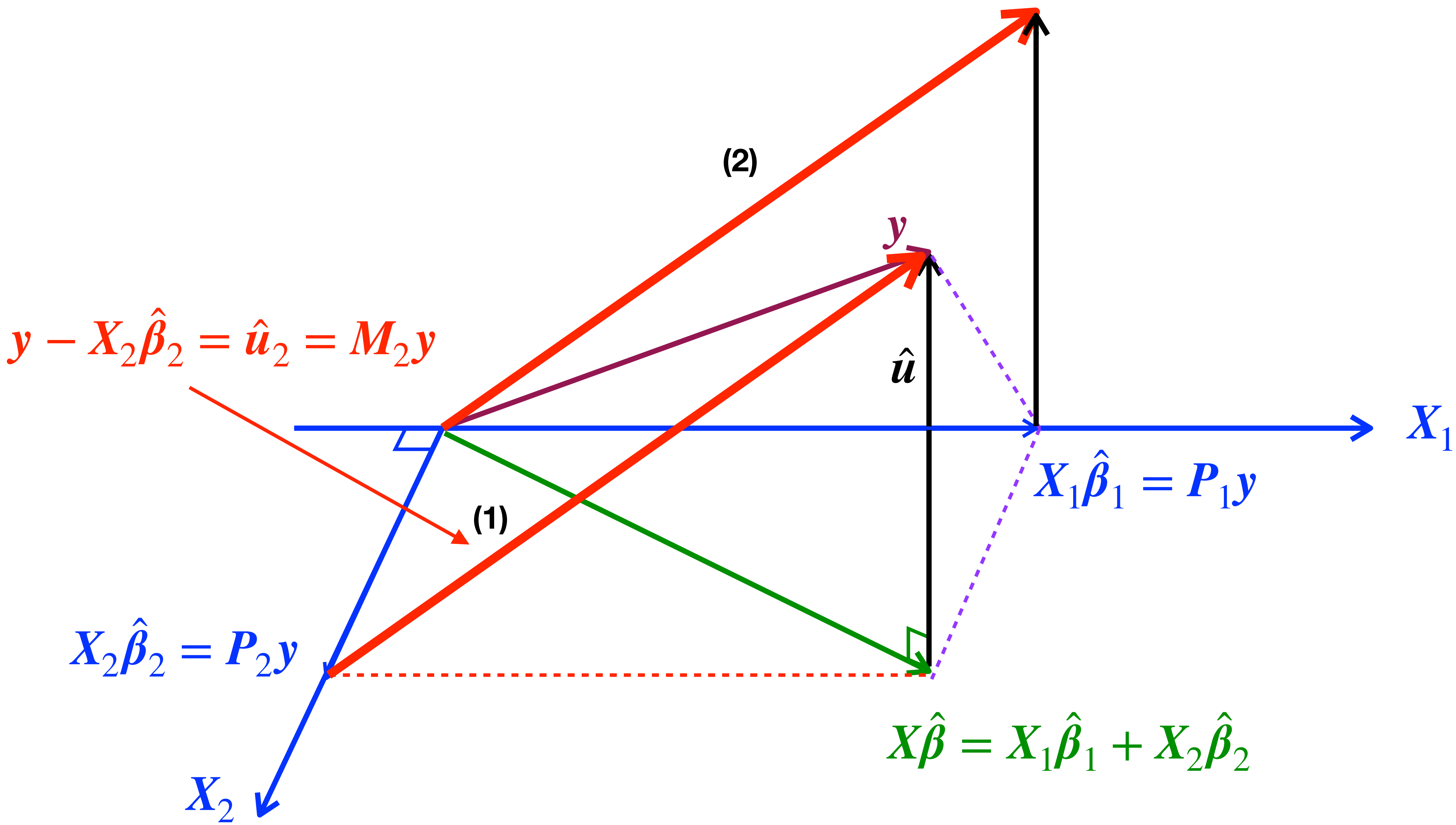
令 $\boldsymbol{P}_1 \equiv \boldsymbol{P}_{\boldsymbol{X}_1} = \boldsymbol{X}_1(\boldsymbol{X}_1^\top\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^\top$

$$\begin{aligned}\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{P}_1 &= \boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{X}_1(\boldsymbol{X}_1^\top\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^\top \\ &= \boldsymbol{X}_1(\boldsymbol{X}_1^\top\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^\top = \boldsymbol{P}_1\end{aligned}$$

因为 $\boldsymbol{X}_1$ 的所有列都包含在 $\mathcal{S}(\boldsymbol{X})$ 中

$$\Rightarrow \quad \begin{aligned}\boldsymbol{P}_1\boldsymbol{y} = \boldsymbol{P}_1\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{y} &= \boldsymbol{P}_1(\boldsymbol{X}_1\hat{\boldsymbol{\beta}}_1 + \boldsymbol{X}_2\hat{\boldsymbol{\beta}}_2) \\ &= \boldsymbol{P}_1\boldsymbol{X}_1\hat{\boldsymbol{\beta}}_1 = \boldsymbol{X}_1\hat{\boldsymbol{\beta}}_1\end{aligned}$$

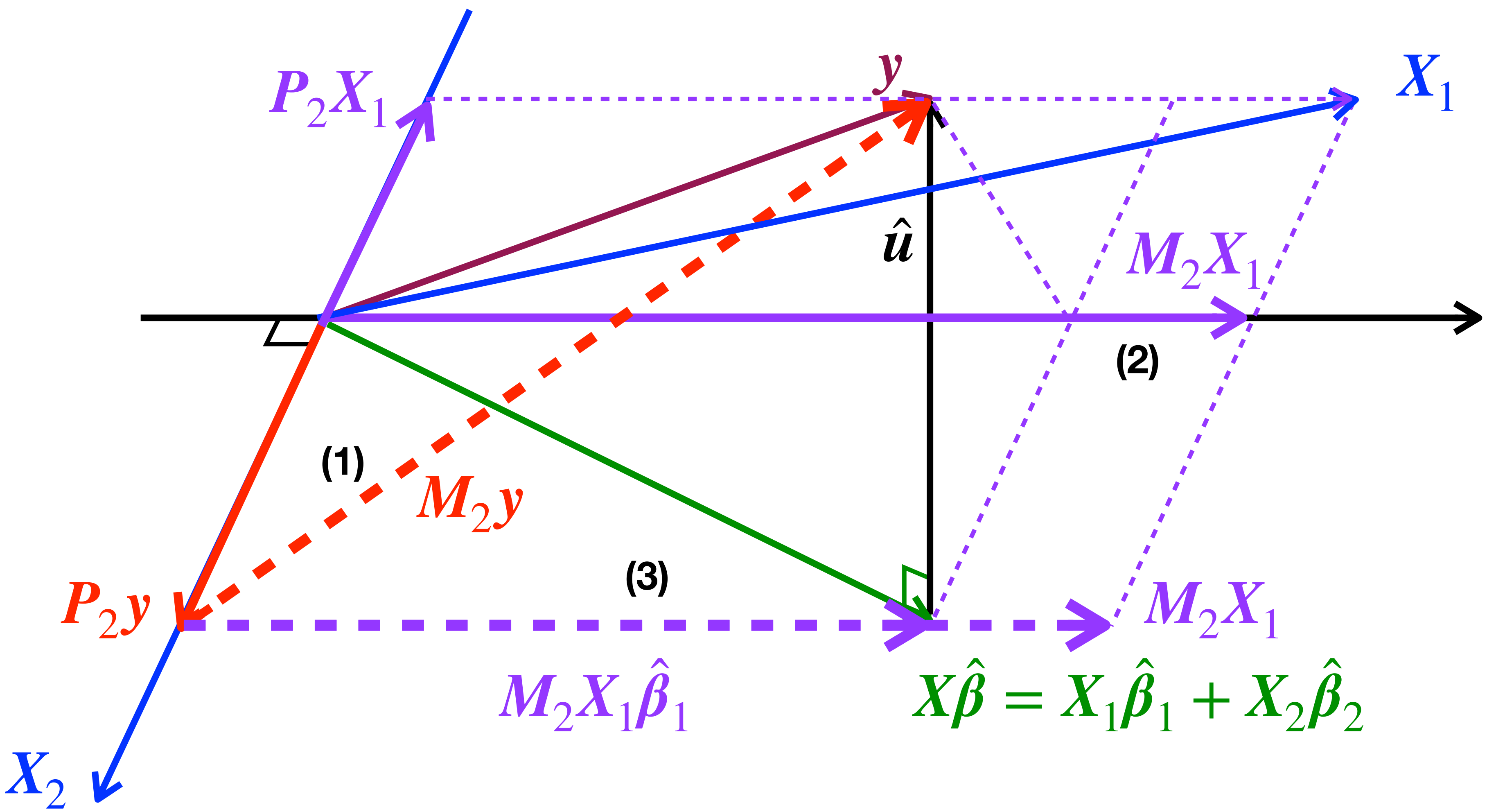
$\boldsymbol{X}_1$ 与 $\boldsymbol{X}_2$ 正交



(1) 拟合 $\boldsymbol{y} = \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{u}_2$ 并得到残差 $\boldsymbol{M}_2\boldsymbol{y}$。

(2) 用 $\boldsymbol{X}_1$ 对 $\boldsymbol{M}_2\boldsymbol{y}$ 进行回归，则得到的参数为 $\hat{\boldsymbol{\beta}}_1$，残差为 $\hat{\boldsymbol{u}}$，与 $\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{u}$ 的拟合结果一致。

当 X_1 与 X_2 非正交



(1) 用 X_2 对 y 进行回归并得到残差 M_2y。

(2) 用 X_2 对 X_1 的每一列进行回归并得到残差 M_2X_1。

(3) 用 M_2X_1 对 M_2y 进行回归，则得到的参数为 $\hat{\beta}_1$，残差为 $\hat{u}$，与 $y = X_1\beta_1 + X_2\beta_2 + u$ 的拟合结果一致。

The Frisch-Waugh-Lovell Theorem

The Frisch-Waugh-Lovell Theorem

考虑下面三个回归模型：

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{u} \qquad (1)$$

$$\boldsymbol{M}_2\boldsymbol{y} = \boldsymbol{M}_2\boldsymbol{X}_1\boldsymbol{b}_1 + \boldsymbol{v} \qquad (2)$$

$$\boldsymbol{y} = \boldsymbol{M}_2\boldsymbol{X}_1\boldsymbol{d}_1 + \boldsymbol{w} \qquad (3)$$

1. 关于 $\boldsymbol{X}_1$ 的系数，模型 (1-3) 的 OLS 估计值相同，即 $\hat{\boldsymbol{\beta}}_1 = \hat{\boldsymbol{b}}_1 = \hat{\boldsymbol{d}}_1$。
2. 模型 (1) 和 (2) 的 OLS 残差相同，和模型 (3) 的残差不同，即 $\hat{\boldsymbol{u}} = \hat{\boldsymbol{v}} \neq \hat{\boldsymbol{w}}$。

- 我们通常称 (1) 为长模型，(2) 和 (3) 为短模型。
- 短模型的作用可以理解为：虽然在长模型中 $\boldsymbol{X}_1$ 与 $\boldsymbol{X}_2$ 非正交，但我们可以想办法找到两个正交矩阵，那就是 $\boldsymbol{M}_2\boldsymbol{X}_1$ 和 $\boldsymbol{X}_2$。同时，$\mathcal{S}(\boldsymbol{X}_1, \boldsymbol{X}_2) = \mathcal{S}(\boldsymbol{M}_2\boldsymbol{X}_1, \boldsymbol{X}_2)$。因此用 $(\boldsymbol{X}_1, \boldsymbol{X}_2)$ 回归 $\boldsymbol{y}$ 得到的参数等同于用 $(\boldsymbol{M}_2\boldsymbol{X}_1, \boldsymbol{X}_2)$ 回归 $\boldsymbol{y}$ 或 $\boldsymbol{M}_2\boldsymbol{y}$ 得到的参数，而后者可以不控制 $\boldsymbol{X}_2$ 。
- 如果样本量很大，则长模型的 OLS 拟合需要用更长的时间和更多的资源去计算 $(\boldsymbol{X}^\top\boldsymbol{X})^{-1}$ 。短模型则可以节省时间和资源。Frisch & Waugh 在 1933 年率先发现了这个结论，当时还没有电子计算机。

FWL 定理的证明

X_1 和 X_2 正交

根据标准方程 $\boldsymbol{X}^\top\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}^\top\boldsymbol{y}$，由模型 $\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{u}$ 可得

$$\begin{bmatrix} \boldsymbol{X}_1^\top \\ \boldsymbol{X}_2^\top \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_1 & \boldsymbol{X}_2 \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_1^\top \\ \boldsymbol{X}_2^\top \end{bmatrix} \boldsymbol{y}$$

$$\Leftrightarrow \quad \begin{bmatrix} \boldsymbol{X}_1^\top\boldsymbol{X}_1 & \boldsymbol{X}_1^\top\boldsymbol{X}_2 \\ \boldsymbol{X}_2^\top\boldsymbol{X}_1 & \boldsymbol{X}_2^\top\boldsymbol{X}_2 \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_1^\top\boldsymbol{y} \\ \boldsymbol{X}_2^\top\boldsymbol{y} \end{bmatrix} \qquad -- (\#)$$

若 $\boldsymbol{X}_1^\top\boldsymbol{X}_2 = \boldsymbol{O}$，则 $\boldsymbol{X}_2^\top\boldsymbol{X}_1 = \boldsymbol{O}$。此时 (#) 可写成

$$\begin{bmatrix} \boldsymbol{X}_1^\top\boldsymbol{X}_1 & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{X}_2^\top\boldsymbol{X}_2 \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_1^\top\boldsymbol{y} \\ \boldsymbol{X}_2^\top\boldsymbol{y} \end{bmatrix} \quad \Rightarrow \quad \begin{bmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2 \end{bmatrix} = \begin{bmatrix} (\boldsymbol{X}_1^\top\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^\top\boldsymbol{y} \\ (\boldsymbol{X}_2^\top\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2^\top\boldsymbol{y} \end{bmatrix}$$

因此，当 $\boldsymbol{X}_1$ 和 $\boldsymbol{X}_2$ 正交时，回归系数 $\hat{\boldsymbol{\beta}}_1$ 和 $\hat{\boldsymbol{\beta}}_2$ 可以分别由两个短模型 $\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{u}_1$ 和 $\boldsymbol{y} = \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{u}_2$ 获得。

FWL 定理的证明

系数的 OLS 估计值相同

一般情况下，由 (#) 可得

$$
\begin{aligned}
\boldsymbol{X}_1^\top \boldsymbol{X}_1 \hat{\boldsymbol{\beta}}_1 + \boldsymbol{X}_1^\top \boldsymbol{X}_2 \hat{\boldsymbol{\beta}}_2 &= \boldsymbol{X}_1^\top \boldsymbol{y} \quad \cdots\cdots \quad (a) \\
\boldsymbol{X}_2^\top \boldsymbol{X}_1 \hat{\boldsymbol{\beta}}_1 + \boldsymbol{X}_2^\top \boldsymbol{X}_2 \hat{\boldsymbol{\beta}}_2 &= \boldsymbol{X}_2^\top \boldsymbol{y} \quad \cdots\cdots \quad (b)
\end{aligned}
$$

由 (b) 可得 $\hat{\boldsymbol{\beta}}_2 = (\boldsymbol{X}_2^\top \boldsymbol{X}_2)^{-1} \boldsymbol{X}_2^\top (\boldsymbol{y} - \boldsymbol{X}_1 \hat{\boldsymbol{\beta}}_1)$。将此式代入 (a) 可得

$$
\begin{aligned}
\boldsymbol{X}_1^\top \boldsymbol{X}_1 \hat{\boldsymbol{\beta}}_1 + \boldsymbol{X}_1^\top \boldsymbol{X}_2 (\boldsymbol{X}_2^\top \boldsymbol{X}_2)^{-1} \boldsymbol{X}_2^\top (\boldsymbol{y} - \boldsymbol{X}_1 \hat{\boldsymbol{\beta}}_1) &= \boldsymbol{X}_1^\top \boldsymbol{y} \\
\left(\boldsymbol{X}_1^\top \boldsymbol{X}_1 - \boldsymbol{X}_1^\top \boldsymbol{X}_2 (\boldsymbol{X}_2^\top \boldsymbol{X}_2)^{-1} \boldsymbol{X}_2^\top \boldsymbol{X}_1 \right) \hat{\boldsymbol{\beta}}_1 &= \left(\boldsymbol{X}_1^\top - \boldsymbol{X}_1^\top \boldsymbol{X}_2 (\boldsymbol{X}_2^\top \boldsymbol{X}_2)^{-1} \boldsymbol{X}_2^\top \right) \boldsymbol{y} \\
\boldsymbol{X}_1^\top \left(\boldsymbol{I} - \boldsymbol{X}_2 (\boldsymbol{X}_2^\top \boldsymbol{X}_2)^{-1} \boldsymbol{X}_2^\top \right) \boldsymbol{X}_1 \hat{\boldsymbol{\beta}}_1 &= \boldsymbol{X}_1^\top \left(\boldsymbol{I} - \boldsymbol{X}_2 (\boldsymbol{X}_2^\top \boldsymbol{X}_2)^{-1} \boldsymbol{X}_2^\top \right) \boldsymbol{y} \\
\boldsymbol{X}_1^\top \boldsymbol{M}_2 \boldsymbol{X}_1 \hat{\boldsymbol{\beta}}_1 &= \boldsymbol{X}_1^\top \boldsymbol{M}_2 \boldsymbol{y}
\end{aligned}
$$

$\boldsymbol{M}_2$ 对称且幂等

$$
\begin{aligned}
\Rightarrow \qquad \hat{\boldsymbol{\beta}}_1 &= (\boldsymbol{X}_1^\top \boldsymbol{M}_2 \boldsymbol{X}_1)^{-1} \boldsymbol{X}_1^\top \boldsymbol{M}_2 \boldsymbol{y} \\
&= (\boldsymbol{X}_1^\top \boldsymbol{M}_2^\top \boldsymbol{M}_2 \boldsymbol{X}_1)^{-1} \boldsymbol{X}_1^\top \boldsymbol{M}_2^\top \boldsymbol{M}_2 \boldsymbol{y} \\
&= \underbrace{\left[(\boldsymbol{M}_2 \boldsymbol{X}_1)^\top \boldsymbol{M}_2 \boldsymbol{X}_1 \right]^{-1} (\boldsymbol{M}_2 \boldsymbol{X}_1)^\top \boldsymbol{M}_2 \boldsymbol{y}}_{\text{模型 (2) 的 OLS 估计 } \hat{\boldsymbol{b}}_1} = \underbrace{\left[(\boldsymbol{M}_2 \boldsymbol{X}_1)^\top \boldsymbol{M}_2 \boldsymbol{X}_1 \right]^{-1} (\boldsymbol{M}_2 \boldsymbol{X}_1)^\top \boldsymbol{y}}_{\text{模型 (3) 的 OLS 估计 } \hat{\boldsymbol{d}}_1}
\end{aligned}
$$

FWL 定理的证明

OLS 残差相同

模型 (1) 的 OLS 残差是

$$\hat{\boldsymbol{u}} = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{y} - \boldsymbol{X}_1\hat{\boldsymbol{\beta}}_1 - \boldsymbol{X}_2\hat{\boldsymbol{\beta}}_2$$

针对最左侧和最右侧的表达式，分别从左边乘以 $\boldsymbol{M}_2$ 可得

$$\boldsymbol{M}_2\hat{\boldsymbol{u}} = \boldsymbol{M}_2\boldsymbol{y} - \boldsymbol{M}_2\boldsymbol{X}_1\hat{\boldsymbol{\beta}}_1 - \boldsymbol{M}_2\boldsymbol{X}_2\hat{\boldsymbol{\beta}}_2$$

$$\hat{\boldsymbol{u}} = \boldsymbol{M}_2\boldsymbol{y} - \boldsymbol{M}_2\boldsymbol{X}_1\hat{\boldsymbol{\beta}}_1$$

$\boldsymbol{M}_2\boldsymbol{X}_2 = \boldsymbol{O}$

$\hat{\boldsymbol{u}} \in \mathcal{S}^{\perp}(\boldsymbol{X}) \subseteq \mathcal{S}^{\perp}(\boldsymbol{X}_2)$

模型 (2) 和 (3) 的残差分别为

$$\hat{\boldsymbol{v}} = \boldsymbol{M}_2\boldsymbol{y} - \boldsymbol{M}_2\boldsymbol{X}_1\hat{\boldsymbol{b}}_1 = \boldsymbol{M}_2\boldsymbol{y} - \boldsymbol{M}_2\boldsymbol{X}_1\hat{\boldsymbol{\beta}}_1$$

$$\hat{\boldsymbol{w}} = \boldsymbol{y} - \boldsymbol{M}_2\boldsymbol{X}_1\hat{\boldsymbol{d}}_1 = \boldsymbol{y} - \boldsymbol{M}_2\boldsymbol{X}_1\hat{\boldsymbol{\beta}}_1$$

因此，一般情况下 $\hat{\boldsymbol{u}} = \hat{\boldsymbol{v}} \neq \hat{\boldsymbol{w}}$，仅当 $\boldsymbol{y} \in \mathcal{S}^{\perp}(\boldsymbol{X}_2)$ 时 $\hat{\boldsymbol{u}} = \hat{\boldsymbol{v}} = \hat{\boldsymbol{w}}$。

如何理解 FWL 定理

- 以模型 (1) 和 (3) 为例：

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{u} \qquad (1)$$

$$\boldsymbol{y} = \boldsymbol{M}_2\boldsymbol{X}_1\boldsymbol{d}_1 + \boldsymbol{w} \qquad (3)$$

- 如果我们的主要研究对象是 $\boldsymbol{X}_1$，则模型 (1) 中的 $\boldsymbol{X}_2$ 有可能和 $\boldsymbol{X}_1$ 存在相关性。

- 模型 (3) 中的 $\boldsymbol{M}_2\boldsymbol{X}_1$ 是用 $\boldsymbol{X}_2$ 解释了 $\boldsymbol{X}_1$ 后剩下的部分，即 $\boldsymbol{X}_1$ 中无法被 $\boldsymbol{X}_2$ 解释的变化。此时的 OLS 估计量 $\hat{\boldsymbol{d}}_1$ 就是 $\boldsymbol{X}_1$ 自身的变动部分对 $\boldsymbol{y}$ 的变动的影响，与 $\boldsymbol{X}_2$ 无关。

- 因此，为了准确估计 (3) 中的 $\hat{\boldsymbol{d}}_1$，我们可以直接估计长模型 (1) 中的系数。也就是所谓的控制了 $\boldsymbol{X}_2$ 中的变量。

FWL 定理的应用

中心化（deviation from the mean, or centering）

这里考虑单变量回归模型

$$\boldsymbol{y} = \beta_1 \boldsymbol{\iota} + \beta_2 \boldsymbol{x} + \boldsymbol{u}$$

$\boldsymbol{x}$ 和 $\boldsymbol{\iota}$ 非正交，但是如果对 $\boldsymbol{x}$ 进行中心化，即定义 $\boldsymbol{z} = \boldsymbol{x} - \bar{x}\boldsymbol{\iota}$，则 $\boldsymbol{z}$ 和 $\boldsymbol{\iota}$ 正交：

$$\boldsymbol{\iota}^\top \boldsymbol{z} = \boldsymbol{\iota}^\top (\boldsymbol{x} - \bar{x}\boldsymbol{\iota}) = \boldsymbol{\iota}^\top \boldsymbol{x} - \boldsymbol{\iota}^\top \bar{x}\boldsymbol{\iota} = n\bar{x} - n\bar{x} = 0$$

又因 $\boldsymbol{M}_\iota = \boldsymbol{I} - \boldsymbol{\iota}(\boldsymbol{\iota}^\top\boldsymbol{\iota})^{-1}\boldsymbol{\iota}^\top = \boldsymbol{I} - \frac{1}{n}\boldsymbol{\iota}\boldsymbol{\iota}^\top$，

$$\boldsymbol{M}_\iota \boldsymbol{x} = \boldsymbol{I}\boldsymbol{x} - \frac{1}{n}\boldsymbol{\iota}\boldsymbol{\iota}^\top \boldsymbol{x} = \boldsymbol{x} - \bar{x}\boldsymbol{\iota} = \boldsymbol{z}$$

可知 $\boldsymbol{x}$ 的中心化可以通过用 $\boldsymbol{\iota}$ 回归 $\boldsymbol{x}$ 的 OLS 残差获得。因此，根据 FWL 定理，

$$\boldsymbol{y} = \beta_1 \boldsymbol{\iota} + \beta_2 \boldsymbol{x} + \boldsymbol{u} \qquad \text{和} \qquad \boldsymbol{M}_\iota \boldsymbol{y} = \boldsymbol{M}_\iota \boldsymbol{x} b + \boldsymbol{v}$$ （对 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 分别进行了中心化处理）

中 $\boldsymbol{x}$ 系数的 OLS 估计值相同，残差相同，即 $\hat{\beta}_2 = \hat{b}$，$\hat{\boldsymbol{u}} = \hat{\boldsymbol{v}}$。

FWL 定理的应用

Adjusting seasonality

时间序列数据的季节性特征一般可以通过加入季节虚拟变量（seasonal dummy）来实现，下面是一个具体模型：

$$\boldsymbol{y}=\delta_1\boldsymbol{s}_1+\delta_2\boldsymbol{s}_2+\delta_3\boldsymbol{s}_3+\delta_4\boldsymbol{s}_4+\boldsymbol{X\beta}+\boldsymbol{u}=\boldsymbol{S\delta}+\boldsymbol{X\beta}+\boldsymbol{u},$$

$$\boldsymbol{s}_1=\begin{bmatrix}1\\0\\0\\0\\1\\\vdots\end{bmatrix},\ \boldsymbol{s}_2=\begin{bmatrix}0\\1\\0\\0\\0\\\vdots\end{bmatrix},\ \boldsymbol{s}_3=\begin{bmatrix}0\\0\\1\\0\\0\\\vdots\end{bmatrix},\ \boldsymbol{s}_4=\begin{bmatrix}0\\0\\0\\1\\0\\\vdots\end{bmatrix}$$

令 $\boldsymbol{M}_S=\boldsymbol{I}-\boldsymbol{S}(\boldsymbol{S}^\top\boldsymbol{S})^{-1}\boldsymbol{S}^\top$，则根据 FWL 定理，可以通过以下模型

$$\boldsymbol{M}_S\boldsymbol{y}=\boldsymbol{M}_S\boldsymbol{X}\boldsymbol{b}+\boldsymbol{v}$$

获得相同的系数估计值和残差。这种先将数据用 $\boldsymbol{M}_S$ 进行投影（用季节虚拟变量进行回归）的操作叫做季节性调整（seasonal adjustment）。FWL 定理告诉我们，进行季节性调整和加入季节虚拟变量的效果是一样的。

一些有用的公式和性质

- 模型 $\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{u}$ 的 OLS 估计可以表达为

$$\hat{\boldsymbol{\beta}}_1 = (\boldsymbol{X}_1^\top \boldsymbol{M}_2 \boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^\top \boldsymbol{M}_2 \boldsymbol{y}$$
$$\hat{\boldsymbol{\beta}}_2 = (\boldsymbol{X}_2^\top \boldsymbol{M}_1 \boldsymbol{X}_2)^{-1}\boldsymbol{X}_2^\top \boldsymbol{M}_1 \boldsymbol{y}$$

- 若 $\boldsymbol{\iota}$ 是解释变量之一，或 $\boldsymbol{\iota} \in \mathcal{S}(\boldsymbol{X})$，则 OLS 残差 $\hat{\boldsymbol{u}}$ 与 $\boldsymbol{\iota}$ 正交，因此残差的均值为零

$$\overline{\hat{u}} = \frac{1}{n}\sum_{t=1}^{n}\hat{u}_t = \frac{1}{n}\boldsymbol{\iota}^\top\hat{\boldsymbol{u}} = 0$$

- OLS 残差 $\hat{\boldsymbol{u}}$ 与每个解释变量 $\boldsymbol{x}_i$ 正交，加上 $\overline{\hat{u}} = 0$，可知 $\hat{\boldsymbol{u}}$ 与每个 $\boldsymbol{x}_i$ 的相关系数都为零

$$r_{\boldsymbol{x}_i,\hat{\boldsymbol{u}}} = \frac{\boldsymbol{x}_i^\top\hat{\boldsymbol{u}} - n\bar{x}_i\overline{\hat{u}}}{\sqrt{\sum_t (x_{ti} - \bar{x}_i)^2 \sum_t (u_t - \overline{\hat{u}})^2}} = 0$$

课外阅读

- Frisch, R. and Waugh, F. V. (1933).
  Partial Time Regressions as Compared with Individual Trends.
  Econometrica, 1:4, 387-410.
  https://www.jstor.org/stable/1907330

- Lovell, M. C. (1963).
  Seasonal Adjustment of Economic Time Series and Multiple Regression Analysis.
  Journal of the American Statistical Association, 58:304, 993-1010.
  https://www.jstor.org/stable/2283327