ECON 452* -- NOTE 7

Dummy Variable Interaction Terms

❑ Model 5: Models with Several Discrete Explanatory Variables

Consider a linear regression model in which **two explanatory variables** are ***discrete* or *categorical* variables**.

To illustrate, suppose the two discrete explanatory variables are ***gender*** and ***industry***.

- ***Gender*** can be represented by means of the following two dummy variables:

  $\mathbf{F_i}$ is a ***female* indicator (dummy) variable**, defined as follows:

  $F_i = 1$ if observation i is female, $= 0$ if observation i is not female.

  $\mathbf{M_i}$ is a ***male* indicator (dummy) variable**, defined as follows:

  $M_i = 1$ if observation i is male, $= 0$ if observation i is not male.

Adding-Up Property of the Gender Indicator Variables $\mathbf{F_i}$ and $\mathbf{M_i}$

$$F_i + M_i = 1 \qquad \forall\ i$$

- ***Industry*** can be represented by means of the following ***industry*** **dummy variables** (assuming a four-level categorization of the variable industry):

  $IN1_i$ = 1 if observation i is in industry 1, = 0 otherwise.

  $IN2_i$ = 1 if observation i is in industry 2, = 0 otherwise.

  $IN3_i$ = 1 if observation i is in industry 3, = 0 otherwise.

  $IN4_i$ = 1 if observation i is in industry 4, = 0 otherwise.

Adding-Up Property of the Industry Indicator Variables:

$$IN1_i + IN2_i + IN3_i + IN4_i = 1 \qquad \forall\ i$$

REVIEW: ***Model 5.2 -- Base Groups for Gender and Industry***

- **Base Groups in Model 5.2**
- ***Males*** are selected as the **base group for *gender***.
- ***Industry 1*** is selected as the **base group for *industry***.

- The **population regression *equation* for Model 5.2** is:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \lambda_f F_i + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i + u_i \qquad \textbf{(5.2)}$$

- The **population regression *function* for Model 5.2** is:

$$E(Y_i \mid X_{i1}, X_{i2}, F_i, IN2_i, IN3_i, IN4_i)$$
$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \lambda_f F_i + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i \qquad (5.2')$$

- The ***female*** **population regression function for Model 5.2** is obtained by setting the female indicator $F_i = 1$ in (5.2'):

$$E(Y_i \mid F_i = 1, X_{i1}, X_{i2}, IN2_i, IN3_i, IN4_i)$$
$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \lambda_f + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i$$
$$= \beta_0 + \lambda_f + \beta_1 X_{i1} + \beta_2 X_{i2} + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i \qquad \textbf{(5.2f)}$$

The female population regression function gives the ***female*** **conditional mean Y** value for *given* values of the regressors X_1, X_2, IN2, IN3, and IN4.

$$E(Y_i \mid X_{i1}, X_{i2}, F_i, IN2_i, IN3_i, IN4_i)$$

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \lambda_f F_i + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i \qquad (5.2')$$

- The ***male*** **population regression function for Model 5.2** is obtained by setting the female indicator $F_i = 0$ in (5.2'):

$$E(Y_i \mid F_i = 0, X_{i1}, X_{i2}, IN2_i, IN3_i, IN4_i)$$

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i \qquad \mathbf{(5.2m)}$$

  The male population regression function gives the ***male*** **conditional mean Y** value for *given* values of the regressors X_1, X_2, IN2, IN3, and IN4.

- Compare the ***female*** **and** ***male*** **population regression functions for Model 5.2**:

  Only the ***intercept coefficient*** **differs** between the male and female regression functions implied by Model 5.2.

  The ***slope coefficients*** **are all identical** in the male and female regression functions for Model 5.2.

  - The **marginal effects** of the ***continuous*** **explanatory variables X_1 and X_2** are ***equal***, or ***identical***, **for *males* and *females***.

  - **Inter-industry differences** in the conditional mean value of Y are ***equal*** **for *males* and *females***. The effects of industry on Y are identical for males and females in Model 5.2.

- The ***female-male difference*** **in conditional mean Y for given values of the regressors** is obtained by subtracting the male population regression function (5.2m) from the female population regression function (5.2f):

  The ***difference*** between the ***female*** **conditional mean Y** for *given* values of the regressors X_1, X_2, IN2, IN3, and IN4 and the ***male*** **conditional mean Y** for the *same* values of the regressors X_1, X_2, IN2, IN3, and IN4 is therefore:

$$
\begin{aligned}
& E(Y_i \mid F_i = 1, x_i^T) - E(Y_i \mid F_i = 0, x_i^T) \\
&\quad = \beta_0 + \lambda_f + \beta_1 X_{i1} + \beta_2 X_{i2} + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i \\
&\qquad - (\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i) \\
&\quad = \beta_0 + \lambda_f + \beta_1 X_{i1} + \beta_2 X_{i2} + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i \\
&\qquad - \beta_0 - \beta_1 X_{i1} - \beta_2 X_{i2} - \pi_2 IN2_i - \pi_3 IN3_i - \pi_4 IN4_i \\
&\quad = \lambda_f \qquad \textbf{(5.2*)}
\end{aligned}
$$

  Note: The ***female-male difference*** **in the conditional mean value of Y** for given values of the regressors X_{i1}, X_{i2}, $IN2_i$, $IN3_i$, and $IN4_i$ is ***a constant***; it does not depend on the value of the regressors X_1 and X_2 or on industry.

Model 5.3 - Version 3 of Model 5: Female-Male Differences in Industry Effects

Allow for ***different* industry effects for *males* and *females*** by introducing into Model 5.2 **three additional regressors** that take the form of ***female interactions* with the three industry indicator variables $IN2_i$, $IN3_i$, and $IN4_i$.**

- The **population regression equation for Model 5.3** can be written as

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \lambda_f F_i + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i + \kappa_2 F_i IN2_i + \kappa_3 F_i IN3_i + \kappa_4 F_i IN4_i + u_i \qquad \textbf{(5.3)}$$

- The **population regression function for Model 5.3** is obtained by taking the conditional expectation of regression equation (5.3) for any given values of the regressors X_{i1}, X_{i2}, F_i, $IN2_i$, $IN3_i$, and $IN4_i$:

$$E(Y_i \mid X_{i1}, X_{i2}, F_i, IN2_i, IN3_i, IN4_i)$$

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \lambda_f F_i + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i + \kappa_2 F_i IN2_i + \kappa_3 F_i IN3_i + \kappa_4 F_i IN4_i \qquad (5.3')$$

- The ***female* population regression function for Model 5.3** is obtained by setting the female indicator $F_i = 1$ in (5.3'):

$$E(Y_i \mid F_i = 1, X_{i1}, X_{i2}, IN2_i, IN3_i, IN4_i)$$

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \lambda_f + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i + \kappa_2 IN2_i + \kappa_3 IN3_i + \kappa_4 IN4_i$$

$$= \beta_0 + \lambda_f + \beta_1 X_{i1} + \beta_2 X_{i2} + (\pi_2 + \kappa_2) IN2_i + (\pi_3 + \kappa_3) IN3_i + (\pi_4 + \kappa_4) IN4_i \qquad \textbf{… (5.3f)}$$

$$E(Y_i \mid X_{i1}, X_{i2}, F_i, IN2_i, IN3_i, IN4_i)$$

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \lambda_f F_i + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i + \kappa_2 F_i IN2_i + \kappa_3 F_i IN3_i + \kappa_4 F_i IN4_i \qquad (5.3')$$

- The ***male*** **population regression function for Model 5.3** is obtained by setting the female indicator $F_i = 0$ in (5.3'):

$$E(Y_i \mid F_i = 0, X_{i1}, X_{i2}, IN2_i, IN3_i, IN4_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i \qquad \textbf{… (5.3m)}$$

- The ***female-male difference*** **in conditional mean Y for given values of the regressors** is obtained by subtracting the male population regression function (5.3m) from the female population regression function (5.3f):

  - The ***female*** **population regression function for Model 5.3** is:

$$E(Y_i \mid F_i = 1, X_{i1}, X_{i2}, IN2_i, IN3_i, IN4_i)$$

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \lambda_f + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i + \kappa_2 IN2_i + \kappa_3 IN3_i + \kappa_4 IN4_i$$

$$= \beta_0 + \lambda_f + \beta_1 X_{i1} + \beta_2 X_{i2} + (\pi_2 + \kappa_2) IN2_i + (\pi_3 + \kappa_3) IN3_i + (\pi_4 + \kappa_4) IN4_i \qquad \textbf{… (5.3f)}$$

  - The ***male*** **population regression function for Model 5.3** is:

$$E(Y_i \mid F_i = 0, X_{i1}, X_{i2}, IN2_i, IN3_i, IN4_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i \qquad \textbf{… (5.3m)}$$

The ***female-male difference*** in the **conditional mean Y** for *given* values of the regressors X_1, X_2, IN2, IN3, and IN4 is:

$$E(Y_i \mid F_i = 1, x_i^T) - E(Y_i \mid F_i = 0, x_i^T)$$

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \lambda_f + (\pi_2 + \kappa_2)IN2_i + (\pi_3 + \kappa_3)IN3_i + (\pi_4 + \kappa_4)IN4_i$$
$$- (\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i)$$

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \lambda_f + (\pi_2 + \kappa_2)IN2_i + (\pi_3 + \kappa_3)IN3_i + (\pi_4 + \kappa_4)IN4_i$$
$$- \beta_0 - \beta_1 X_{i1} - \beta_2 X_{i2} - \pi_2 IN2_i - \pi_3 IN3_i - \pi_4 IN4_i$$

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \lambda_f + \pi_2 IN2_i + \kappa_2 IN2_i + \pi_3 IN3_i + \kappa_3 IN3_i + \pi_4 IN4_i + \kappa_4 IN4_i$$
$$- \beta_0 - \beta_1 X_{i1} - \beta_2 X_{i2} - \pi_2 IN2_i - \pi_3 IN3_i - \pi_4 IN4_i$$

$$= \lambda_f + \kappa_2 IN2_i + \kappa_3 IN3_i + \kappa_4 IN4_i \qquad \textbf{(5.3*)}$$

- The ***female* population regression function for Model 5.3** is:

$$E(Y_i \mid F_i = 1, X_{i1}, X_{i2}, IN2_i, IN3_i, IN4_i)$$

$$= (\beta_0 + \lambda_f) + \beta_1 X_{i1} + \beta_2 X_{i2} + (\pi_2 + \kappa_2)IN2_i + (\pi_3 + \kappa_3)IN3_i + (\pi_4 + \kappa_4)IN4_i \qquad \text{... (5.3f)}$$

The ***female* population regression function for Model 5.3** implies that the conditional mean value of Y for females differs across industries:

1. The **conditional mean value of Y** for **females in *industry 1*** is:

$$E(Y_i \mid F_i = 1, X_{i1}, X_{i2}, IN2_i = 0, IN3 = 0_i, IN4_i = 0) = (\beta_0 + \lambda_f) + \beta_1 X_{i1} + \beta_2 X_{i2}$$

2. The **conditional mean value of Y** for **females in *industry 2*** is:

$$E(Y_i \mid F_i = 1, X_{i1}, X_{i2}, IN2_i = 1, IN3 = 0_i, IN4_i = 0) = (\beta_0 + \lambda_f) + \beta_1 X_{i1} + \beta_2 X_{i2} + (\pi_2 + \kappa_2)$$

3. The **conditional mean value of Y** for **females in *industry 3*** is:

$$E(Y_i \mid F_i = 1, X_{i1}, X_{i2}, IN2_i = 0, IN3 = 1_i, IN4_i = 0) = (\beta_0 + \lambda_f) + \beta_1 X_{i1} + \beta_2 X_{i2} + (\pi_3 + \kappa_3)$$

4. The **conditional mean value of Y** for **females in *industry 4*** is:

$$E(Y_i \mid F_i = 1, X_{i1}, X_{i2}, IN2_i = 0, IN3 = 0_i, IN4_i = 1) = (\beta_0 + \lambda_f) + \beta_1 X_{i1} + \beta_2 X_{i2} + (\pi_4 + \kappa_4)$$

$$E(Y_i \mid X_{i1}, X_{i2}, F_i, IN2_i, IN3_i, IN4_i)$$

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \lambda_f F_i + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i + \kappa_2 F_i IN2_i + \kappa_3 F_i IN3_i + \kappa_4 F_i IN4_i \qquad (5.3')$$

- The ***male* population regression function for Model 5.3** is:

$$E(Y_i \mid F_i = 0, X_{i1}, X_{i2}, IN2_i, IN3_i, IN4_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i \qquad \textbf{... (5.3m)}$$

The ***male* population regression function for Model 5.3** implies that the conditional mean value of Y for males differs across industries:

1. The **conditional mean value of Y** for **males in *industry 1*** is:

$$E(Y_i \mid F_i = 0, X_{i1}, X_{i2}, IN2_i = 0, IN3 = 0_i, IN4_i = 0) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2}$$

2. The **conditional mean value of Y** for **males in *industry 2*** is:

$$E(Y_i \mid F_i = 0, X_{i1}, X_{i2}, IN2_i = 1, IN3 = 0_i, IN4_i = 0) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \pi_2$$

3. The **conditional mean value of Y** for **males in *industry 3*** is:

$$E(Y_i \mid F_i = 0, X_{i1}, X_{i2}, IN2_i = 0, IN3 = 1_i, IN4_i = 0) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \pi_3$$

4. The **conditional mean value of Y** for **males in *industry 4*** is:

$$E(Y_i \mid F_i = 0, X_{i1}, X_{i2}, IN2_i = 0, IN3 = 0_i, IN4_i = 1) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \pi_4$$

The ***difference*** between the ***female*** **conditional mean Y** for *given* values of the regressors X_1, X_2, IN2, IN3, and IN4 and the ***male*** **conditional mean Y** for the *same* values of the regressors X_1, X_2, IN2, IN3, and IN4 is:

$$E(Y_i \mid F_i = 1, x_i^T) - E(Y_i \mid F_i = 0, x_i^T) = \lambda_f + \kappa_2 IN2_i + \kappa_3 IN3_i + \kappa_4 IN4_i \qquad \textbf{(5.3*)}$$

1. The ***female-male difference*** in **conditional mean Y for *industry 1*** for given values of X_1 and X_2 is obtained by setting $IN2_i = 0$ and $IN3_i = 0$ and $IN4_i = 0$ in (5.3*):

$$E(Y_i \mid F_i = 1, IN2_i = 0, IN3_i = 0, IN4_i = 0) - E(Y_i \mid F_i = 0, IN2_i = 0, IN3_i = 0, IN4_i = 0) = \lambda_f$$

2. The ***female-male difference*** in **conditional mean Y for *industry 2*** for given values of X_1 and X_2 is obtained by setting $IN2_i = 1$ and $IN3_i = 0$ and $IN4_i = 0$ in (5.3*):

$$E(Y_i \mid F_i = 1, IN2_i = 1, IN3_i = 0, IN4_i = 0) - E(Y_i \mid F_i = 0, IN2_i = 1, IN3_i = 0, IN4_i = 0) = \lambda_f + \kappa_2$$

3. The ***female-male difference*** in **conditional mean Y for *industry 3*** for given values of X_1 and X_2 is obtained by setting $IN2_i = 0$ and $IN3_i = 1$ and $IN4_i = 0$ in (5.3*):

$$E(Y_i \mid F_i = 1, IN2_i = 0, IN3_i = 1, IN4_i = 0) - E(Y_i \mid F_i = 0, IN2_i = 0, IN3_i = 1, IN4_i = 0) = \lambda_f + \kappa_3$$

4. The ***female-male difference*** in **conditional mean Y for *industry 4*** for given values of X_1 and X_2 is obtained by setting $IN2_i = 0$ and $IN3_i = 0$ and $IN4_i = 1$ in (5.3*):

$$E(Y_i \mid F_i = 1, IN2_i = 0, IN3_i = 0, IN4_i = 1) - E(Y_i \mid F_i = 0, IN2_i = 0, IN3_i = 0, IN4_i = 1) = \lambda_f + \kappa_4$$

Propositions to Test in Model 5.3

$$E(Y_i \mid F_i = 1, x_i^T) - E(Y_i \mid F_i = 0, x_i^T) = \lambda_f + \kappa_2 IN2_i + \kappa_3 IN3_i + \kappa_4 IN4_i \qquad \textbf{(5.3*)}$$

<u>Test 1</u>: The ***female-male difference* in conditional mean Y *equals zero* for all observations**.

H_0: $\lambda_f = 0 \text{ and } \kappa_2 = 0 \text{ and } \kappa_3 = 0 \text{ and } \kappa_4 = 0$
H_1: $\lambda_f \neq 0 \text{ and/or } \kappa_2 \neq 0 \text{ and/or } \kappa_3 \neq 0 \text{ and/or } \kappa_4 \neq 0$

<u>Test 2</u>: The ***female-male difference* in conditional mean Y *equals a constant, i.e., does not depend on industry***:

H_0: $\kappa_2 = 0 \text{ and } \kappa_3 = 0 \text{ and } \kappa_4 = 0$
H_1: $\kappa_2 \neq 0 \text{ and/or } \kappa_3 \neq 0 \text{ and/or } \kappa_4 \neq 0$

- ***Interpretation of the coefficients in Model 5.3***

$$E(Y_i \mid X_{i1}, X_{i2}, F_i, IN2_i, IN3_i, IN4_i)$$

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \lambda_f F_i + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i + \kappa_2 F_i IN2_i + \kappa_3 F_i IN3_i + \kappa_4 F_i IN4_i \qquad (5.3')$$

β_0 = intercept for ***males*** **in industry 1**

$\beta_0 + \lambda_f$ = intercept for ***females*** **in industry 1**

λ_f = ***female*** **industry 1** intercept – ***male*** **industry 1** intercept

$\beta_0 + \pi_2$ = intercept for ***males*** **in industry 2**

$\beta_0 + \lambda_f + \pi_2 + \kappa_2$ = intercept for ***females*** **in industry 2**

$\lambda_f + \kappa_2$ = ***female*** **industry 2** intercept – ***male*** **industry 2** intercept

$\beta_0 + \pi_3$ = intercept for ***males*** **in industry 3**

$\beta_0 + \lambda_f + \pi_3 + \kappa_3$ = intercept for ***females*** **in industry 3**

$\lambda_f + \kappa_3$ = ***female*** **industry 3** intercept – ***male*** **industry 3** intercept

$\beta_0 + \pi_4$ = intercept for ***males*** **in industry 4**

$\beta_0 + \lambda_f + \pi_4 + \kappa_4$ = intercept for ***females*** **in industry 4**

$\lambda_f + \kappa_4$ = ***female*** **industry 4** intercept – ***male*** **industry 4** intercept

- ***Interpretation of the coefficients in Model 5.3 (continued)***

$$E(Y_i \mid X_{i1}, X_{i2}, F_i, IN2_i, IN3_i, IN4_i)$$

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \lambda_f F_i + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i + \kappa_2 F_i IN2_i + \kappa_3 F_i IN3_i + \kappa_4 F_i IN4_i \qquad (5.3')$$

π_2 = ***male*** **industry 2** intercept – ***male*** **industry 1** intercept
π_3 = ***male*** **industry 3** intercept – ***male*** **industry 1** intercept
π_4 = ***male*** **industry 4** intercept – ***male*** **industry 1** intercept

$\pi_2 + \kappa_2$ = ***female*** **industry 2** intercept – ***female*** **industry 1** intercept
$\pi_3 + \kappa_3$ = ***female*** **industry 3** intercept – ***female*** **industry 1** intercept
$\pi_4 + \kappa_4$ = ***female*** **industry 4** intercept – ***female*** **industry 1** intercept

What are the κ_j coefficients in Model 5.3 (j = 2, 3, 4)?

κ_2 = ***female* industry 2** intercept – ***female* industry 1** intercept
– (***male* industry 2** intercept – ***male* industry 1** intercept)

= female industry 2 effect – male industry 2 effect (relative to industry 1)

= female-male difference in intercepts for industry 2
– female-male difference in intercepts for industry 1

κ_3 = ***female* industry 3** intercept – ***female* industry 1** intercept
– (***male* industry 3** intercept – ***male* industry 1** intercept)

= female industry 3 effect – male industry 3 effect (relative to industry 1)

= female-male difference in intercepts for industry 3
– female-male difference in intercepts for industry 1

κ_4 = ***female* industry 4** intercept – ***female* industry 1** intercept
– (***male* industry 4** intercept – ***male* industry 1** intercept)

= female industry 4 effect – male industry 4 effect (relative to industry 1)

= female-male difference in intercepts for industry 4
– female-male difference in intercepts for industry 1

Difference-in-differences interpretation of κ_2 coefficient in Model 5.3

Intercept Coefficients for Females and Males in Industries 1 and 2 - Model 5.3

	1 **Ind 2** ($IN2_i = 1$)	**2** **Ind 1** ($IN1_i = 1$)	***Col. 1 – Col. 2***
1. Females ($F_i = 1$)	$\beta_0 + \lambda_f + \pi_2 + \kappa_2$	$\beta_0 + \lambda_f$	$\pi_2 + \kappa_2$
2. Males ($F_i = 0$)	$\beta_0 + \pi_2$	β_0	π_2
Row 1 – Row 2	$\lambda_f + \kappa_2$	λ_f	κ_2

Interpretation 1: within each column, subtract the element in row 2 from the element in row 1

$\lambda_f + \kappa_2$ = ***Female-Male*** **difference** in intercepts for ***Industry 2***

λ_f = ***Female-Male*** **difference** in intercepts for ***Industry 1***

κ_2 = ***Female-Male*** **difference** in intercepts for ***Industry 2***
minus
Female-Male **difference** in intercepts for ***Industry 1***

Difference-in-differences interpretation of κ_2 coefficient in Model 5.3

Intercept Coefficients for Females and Males in Industries 1 and 2 - Model 5.3

	1 **Ind 2** ($IN2_i = 1$)	**2** **Ind 1** ($IN1_i = 1$)	***Col. 1 – Col. 2***
1. Females ($F_i = 1$)	$\beta_0 + \lambda_f + \pi_2 + \kappa_2$	$\beta_0 + \lambda_f$	$\pi_2 + \kappa_2$
2. Males ($F_i = 0$)	$\beta_0 + \pi_2$	β_0	π_2
Row 1 – Row 2	$\lambda_f + \kappa_2$	λ_f	κ_2

Interpretation 2: within each row, subtract the element in column 2 from the element in column 1

$\pi_2 + \kappa_2$ = ***Industry 2-Industry 1*** **difference** in intercepts for ***Females***

π_2 = ***Industry 2-Industry 1*** **difference** in intercepts for ***Males***

κ_2 = ***Industry 2-Industry 1*** **difference** in intercepts for ***Females***
minus
Industry 2-Industry 1 **difference** in intercepts for ***Males***

Two interpretations of the coefficient κ_2 in Model 5.3

The **population regression function for Model 5.3** is:

$$E(Y_i \mid X_{i1}, X_{i2}, F_i, IN2_i, IN3_i, IN4_i)$$
$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \lambda_f F_i + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i + \kappa_2 F_i IN2_i + \kappa_3 F_i IN3_i + \kappa_4 F_i IN4_i \qquad (5.3')$$

Both interpretations use the **conditional mean values of Y** for **<u>four</u> subgroups**:

1. Conditional mean Y for ***females* in *industry 2***: in regression function (5.3'), set $F_i = 1$, $IN2_i = 1$, $IN3_i = 0$, $IN4_i = 0$

   $$E(Y_i \mid X_{i1}, X_{i2}, F_i = 1, IN2_i = 1, IN3_i = 0, IN4_i = 0) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \lambda_f + \pi_2 + \kappa_2$$

2. Conditional mean Y for ***males* in *industry 2***: in regression function (5.3'), set $F_i = 0$, $IN2_i = 1$, $IN3_i = 0$, $IN4_i = 0$

   $$E(Y_i \mid X_{i1}, X_{i2}, F_i = 0, IN2_i = 1, IN3_i = 0, IN4_i = 0) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \pi_2$$

$$E(Y_i \mid X_{i1}, X_{i2}, F_i, IN2_i, IN3_i, IN4_i)$$
$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \lambda_f F_i + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i + \kappa_2 F_i IN2_i + \kappa_3 F_i IN3_i + \kappa_4 F_i IN4_i \qquad (5.3')$$

3. Conditional mean Y for ***females* in *industry 1***: in regression function (5.3'), set $F_i = 1$, $IN2_i = 0$, $IN3_i = 0$, $IN4_i = 0$

$$E(Y_i \mid X_{i1}, X_{i2}, F_i = 1, IN2_i = 0, IN3_i = 0, IN4_i = 0) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \lambda_f$$

4. Conditional mean Y for ***males* in *industry 1***: in regression function (5.3'), set $F_i = 0$, $IN2_i = 0$, $IN3_i = 0$, $IN4_i = 0$

$$E(Y_i \mid X_{i1}, X_{i2}, F_i = 0, IN2_i = 0, IN3_i = 0, IN4_i = 0) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2}$$

<u>Interpretation 1</u> of κ_2 in Model 5.3

(1) Female-male difference in conditional mean Y **for *industry 2***:

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \lambda_f + \pi_2 + \kappa_2 - \beta_0 - \beta_1 X_{i1} - \beta_2 X_{i2} - \pi_2$$

$$= \lambda_f + \kappa_2$$

(2) Female-male difference in conditional mean Y **for *industry 1***:

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \lambda_f - \beta_0 - \beta_1 X_{i1} - \beta_2 X_{i2}$$

$$= \lambda_f$$

Subtract (2) from (1):

Difference-in-differences* = female-male difference** in conditional mean Y **for *industry 2 <u>minus</u>* female-male difference** in conditional mean Y **for *industry 1:

$$= \lambda_f + \kappa_2 - \lambda_f = \kappa_2$$

<u>Interpretation 2</u> of κ_2 in Model 5.3

(1) Industry 2-industry 1 difference in conditional mean Y for ***females***:

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \lambda_f + \pi_2 + \kappa_2$$
$$- \beta_0 - \beta_1 X_{i1} - \beta_2 X_{i2} - \lambda_f$$

$$= \pi_2 + \kappa_2$$

(2) Industry 2-industry 1 difference in conditional mean Y for ***males***:

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \pi_2 - \beta_0 - \beta_1 X_{i1} - \beta_2 X_{i2} = \pi_2$$

Subtract (2) from (1):

Difference-in-differences* = industry 2-industry 1 difference** in conditional mean Y **for *females* <u>*minus*</u> industry 2-industry 1 difference** in conditional mean Y **for *males:

$$= \pi_2 + \kappa_2 - \pi_2 = \kappa_2$$

Tests for Industry Effects in Model 5.3

<u>Test 1</u>: Test for Industry Effects for *Females* in Model 5.3

- The **population regression function for Model 5.3** gives the conditional mean value of Y for given values of the regressors X_{i1}, X_{i2}, F_i, $IN2_i$, $IN3_i$, and $IN4_i$:

$$E(Y_i \mid X_{i1}, X_{i2}, F_i, IN2_i, IN3_i, IN4_i)$$

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \lambda_f F_i + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i + \kappa_2 F_i IN2_i + \kappa_3 F_i IN3_i + \kappa_4 F_i IN4_i \qquad (5.3')$$

- The ***female*** **population regression function for Model 5.3** is obtained by setting the female indicator $F_i = 1$ in (5.3'):

$$E(Y_i \mid F_i = 1, X_{i1}, X_{i2}, IN2_i, IN3_i, IN4_i)$$

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \lambda_f + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i + \kappa_2 IN2_i + \kappa_3 IN3_i + \kappa_4 IN4_i$$

$$= \beta_0 + \lambda_f + \beta_1 X_{i1} + \beta_2 X_{i2} + (\pi_2 + \kappa_2) IN2_i + (\pi_3 + \kappa_3) IN3_i + (\pi_4 + \kappa_4) IN4_i \qquad \textbf{… (5.3f)}$$

Hypothesis Test: Test the proposition of **no industry effects for *females*** – i.e., the proposition that there are no inter-industry differences in conditional mean Y values for females.

H_0: $\pi_2 + \kappa_2 = 0$ *and* $\pi_3 + \kappa_3 = 0$ *and* $\pi_4 + \kappa_4 = 0$

H_1: $\pi_2 + \kappa_2 \neq 0$ *and/or* $\pi_3 + \kappa_3 \neq 0$ *and/or* $\pi_4 + \kappa_4 \neq 0$

Restricted Model for Females: Substitute the three restrictions specified by H_0 into the unrestricted female regression function (5.3f) to get the restricted female regression function.

$$E(Y_i \mid F_i = 1, X_{i1}, X_{i2}, IN2_i, IN3_i, IN4_i)$$

$$= \beta_0 + \lambda_f + \beta_1 X_{i1} + \beta_2 X_{i2} + (\pi_2 + \kappa_2) IN2_i + (\pi_3 + \kappa_3) IN3_i + (\pi_4 + \kappa_4) IN4_i$$

$$= \beta_0 + \lambda_f + \beta_1 X_{i1} + \beta_2 X_{i2}$$

<u>Test 2</u>: Test for Industry Effects for *Males* in Model 5.3

- The ***male* population regression function for Model 5.3** is obtained by setting the female indicator $F_i = 0$ in (5.3'):

$$E(Y_i \mid X_{i1}, X_{i2}, F_i, IN2_i, IN3_i, IN4_i)$$

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \lambda_f F_i + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i + \kappa_2 F_i IN2_i + \kappa_3 F_i IN3_i + \kappa_4 F_i IN4_i \quad (5.3')$$

$$E(Y_i \mid F_i = 0, X_{i1}, X_{i2}, IN2_i, IN3_i, IN4_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i \quad \textbf{(5.3m)}$$

Hypothesis Test: Test the proposition of **no industry effects for *males*** – i.e., the proposition that there are no inter-industry differences in conditional mean Y values for males.

H_0: $\pi_2 = 0$ *and* $\pi_3 = 0$ *and* $\pi_4 = 0$

H_1: $\pi_2 \neq 0$ *and/or* $\pi_3 \neq 0$ *and/or* $\pi_4 \neq 0$

Restricted Model for Males: Substitute the three restrictions specified by H_0 into the unrestricted male regression function (5.3f) to get the ***restricted* male regression function**.

$$E(Y_i \mid F_i = 0, X_{i1}, X_{i2}, IN2_i, IN3_i, IN4_i)$$

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i \quad \textbf{... (5.3m)}$$

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2}$$

<u>Test 3</u>: Test for *Female-Male Differences* in Industry Effects in Model 5.3

Proposition: Test the proposition of ***equal* industry effects for *females* and *males*** – i.e., the proposition that the inter-industry differences in conditional mean Y values for females are identical to the inter-industry differences in conditional mean Y values for males.

Null and Alternative Hypotheses:

H_0: $\kappa_2 = 0$ *and* $\kappa_3 = 0$ *and* $\kappa_4 = 0$

H_1: $\kappa_2 \neq 0$ *and/or* $\kappa_3 \neq 0$ *and/or* $\kappa_4 \neq 0$

Restricted Model: Substitute the three restrictions specified by H_0 into the unrestricted pooled regression function (5.3') for Model 5.3.

$$E(Y_i \mid X_{i1}, X_{i2}, F_i, IN2_i, IN3_i, IN4_i)$$

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \lambda_f F_i + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i + \kappa_2 F_i IN2_i + \kappa_3 F_i IN3_i + \kappa_4 F_i IN4_i \qquad (5.3')$$

The ***restricted* pooled regression function** under H_0 is therefore:

$$E(Y_i \mid X_{i1}, X_{i2}, F_i, IN2_i, IN3_i, IN4_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \lambda_f F_i + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i$$

<u>Note</u>: This restricted model allows only for an intercept coefficient difference between males and females.

Pooling Male and Female Regression Functions with *Different* Regressor Sets – Model 5.3

- ***Question:*** How can we formulate a ***restricted*** version of the pooled **Model 5.3** when ***male* and *female* regression functions have *different* regressor sets?**

  Consider the following test outcomes for the previous three hypothesis tests:

  Hypothesis Test 1: Test the proposition of **no industry effects for *females***.

  H_0: $\pi_2 + \kappa_2 = 0$ *and* $\pi_3 + \kappa_3 = 0$ *and* $\pi_4 + \kappa_4 = 0$ we **retain** this H_0

  Hypothesis Test 2: Test the proposition of **no industry effects for *males***.

  H_0: $\pi_2 = 0$ *and* $\pi_3 = 0$ *and* $\pi_4 = 0$ we **reject** this H_0

  Hypothesis Test 3: Test the proposition of **no female-male *differences* in industry effects**.

  H_0: $\kappa_2 = 0$ *and* $\kappa_3 = 0$ *and* $\kappa_4 = 0$ we **reject** this H_0

- ***Question:*** What is the ***restricted*** version of the pooled **Model 5.3** implied by this set of three test outcomes?

- ***Derivation of Restricted Pooled Model 5.3*** that incorporates industry effects for males but not for females.

1. Write the **unrestricted male** regression equation ***with*** **industry effects**:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i + u_i \qquad \textbf{... (5.3m)}$$

2. Write **restricted female** regression equation **with *no* industry effects**:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \lambda_f + u_i$$

3. Multiply the male regression equation in step 1 by the male indicator variable M_i, and the female regression equation in step 2 by the female indicator variable F_i:

$$M_i Y_i = \beta_0 M_i + \beta_1 M_i X_{i1} + \beta_2 M_i X_{i2} + \pi_2 M_i IN2_i + \pi_3 M_i IN3_i + \pi_4 M_i IN4_i + M_i u_i$$

$$F_i Y_i = \beta_0 F_i + \beta_1 F_i X_{i1} + \beta_2 F_i X_{i2} + \lambda_f F_i + F_i u_i$$

4. **Add** the above regression equations in step 3 to obtain the corresponding pooled regression equation:

$$\begin{aligned} M_i Y_i + F_i Y_i = {} & \beta_0 M_i + \beta_1 M_i X_{i1} + \beta_2 M_i X_{i2} + \pi_2 M_i IN2_i + \pi_3 M_i IN3_i + \pi_4 M_i IN4_i \\ & + \beta_0 F_i + \beta_1 F_i X_{i1} + \beta_2 F_i X_{i2} + \lambda_f F_i + M_i u_i + F_i u_i \end{aligned}$$

$$M_iY_i + F_iY_i = \beta_0 M_i + \beta_1 M_i X_{i1} + \beta_2 M_i X_{i2} + \pi_2 M_i IN2_i + \pi_3 M_i IN3_i + \pi_4 M_i IN4_i$$
$$+ \beta_0 F_i + \beta_1 F_i X_{i1} + \beta_2 F_i X_{i2} + \lambda_f F_i + M_i u_i + F_i u_i$$

5. **Collect** like terms in the regression coefficients in the above regression equation in step 4:

$$(M_i + F_i)Y_i = \beta_0 (M_i + F_i) + \beta_1 (M_i + F_i) X_{i1} + \beta_2 (M_i + F_i) X_{i2}$$
$$+ \pi_2 M_i IN2_i + \pi_3 M_i IN3_i + \pi_4 M_i IN4_i + \lambda_f F_i + (M_i + F_i) u_i$$

6. Use the **adding-up property** $M_i + F_i = 1$ for all i to simplify the pooled regression equation in step 5:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \pi_2 M_i IN2_i + \pi_3 M_i IN3_i + \pi_4 M_i IN4_i + \lambda_f F_i + u_i$$

- ***Result:*** This pooled regression equation allows for ***different* male and female *intercept* coefficients**, and includes ***industry effects*** **<u>only</u> in the *male* regression function**.

- ***Analysis:*** The pooled regression equation that allows for different male and female intercept coefficients but includes industry effects only for males is:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \pi_2 M_i IN2_i + \pi_3 M_i IN3_i + \pi_4 M_i IN4_i + \lambda_f F_i + u_i$$

The **pooled regression function** for this pooled regression equation is:

$$E(Y_i \,|\, X_{i1}, X_{i2}, M_i, F_i, IN2_i, IN3_i, IN4_i)$$
$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \pi_2 M_i IN2_i + \pi_3 M_i IN3_i + \pi_4 M_i IN4_i + \lambda_f F_i$$

  - The ***male* regression function** is obtained by setting $M_i = 1$ and $F_i = 0$ in the above pooled regression function:

$$E(Y_i \,|\, X_{i1}, X_{i2}, M_i = 1, F_i = 0, IN2_i, IN3_i, IN4_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i$$

  - The ***female* regression function** is obtained by setting $M_i = 0$ and $F_i = 1$ in the above pooled regression function:

$$E(Y_i \,|\, X_{i1}, X_{i2}, M_i = 0, F_i = 1, IN2_i, IN3_i, IN4_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \lambda_f$$
$$= \beta_0 + \lambda_f + \beta_1 X_{i1} + \beta_2 X_{i2}$$

- The ***female-male difference*** **in conditional mean Y for given values of the regressors** is obtained by subtracting the male population regression function from the female population regression function.

The ***female*** **regression function** is:

$$E(Y_i \mid X_{i1}, X_{i2}, M_i = 0, F_i = 1, IN2_i, IN3_i, IN4_i) = \beta_0 + \lambda_f + \beta_1 X_{i1} + \beta_2 X_{i2}$$

The ***male*** **regression function** is:

$$E(Y_i \mid X_{i1}, X_{i2}, M_i = 1, F_i = 0, IN2_i, IN3_i, IN4_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i$$

The ***female-male difference*** in **conditional mean Y** for given values of the regressors X_1, X_2, IN2, IN3, and IN4 is therefore:

$$E(Y_i \mid X_{i1}, X_{i2}, M_i = 0, F_i = 1, IN2_i, IN3_i, IN4_i)$$

$$- \; E(Y_i \mid X_{i1}, X_{i2}, M_i = 1, F_i = 0, IN2_i, IN3_i, IN4_i)$$

$$= \beta_0 + \lambda_f + \beta_1 X_{i1} + \beta_2 X_{i2} - \beta_0 - \beta_1 X_{i1} - \beta_2 X_{i2} - \pi_2 IN2_i - \pi_3 IN3_i - \pi_4 IN4_i$$

$$= \lambda_f - \pi_2 IN2_i - \pi_3 IN3_i - \pi_4 IN4_i$$

- The ***female-male difference*** in **conditional mean Y** for given values of the regressors X_1 and X_2 is different for each of the four industries:

$$E(Y_i \mid X_{i1}, X_{i2}, M_i = 0, F_i = 1, IN2_i, IN3_i, IN4_i) - E(Y_i \mid X_{i1}, X_{i2}, M_i = 1, F_i = 0, IN2_i, IN3_i, IN4_i)$$

$$= \lambda_f - \pi_2 IN2_i - \pi_3 IN3_i - \pi_4 IN4_i$$

$$= \lambda_f \qquad \text{for } \textit{Industry 1} \; (IN1_i = 1;\ IN2_i = IN3_i = IN4_i = 0)$$

$$= \lambda_f - \pi_2 \qquad \text{for } \textit{Industry 2} \; (IN2_i = 1;\ IN1_i = IN3_i = IN4_i = 0)$$

$$= \lambda_f - \pi_3 \qquad \text{for } \textit{Industry 3} \; (IN3_i = 1;\ IN1_i = IN2_i = IN4_i = 0)$$

$$= \lambda_f - \pi_4 \qquad \text{for } \textit{Industry 4} \; (IN4_i = 1;\ IN1_i = IN2_i = IN3_i = 0)$$

<u>Model 5.4</u>: Female-Male Differences in Industry Effects and in Marginal Effects of X_1 and X_2

Add to Model 5.3 female-male differences in the marginal effects of the two ***continuous*** **explanatory variables X_1 and X_2**.

Allow the effects of the regressors X_1 and X_2 to differ between males and females by introducing into Model 5.3 **two additional regressors** that take the form of ***female interactions*** **with the two regressors X_{i1} and X_{i2}**.

- The **population regression equation for Model 5.4** can be written as

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \lambda_f F_i + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i + \eta_1 F_i X_{i1} + \eta_2 F_i X_{i2} + \kappa_2 F_i IN2_i + \kappa_3 F_i IN3_i + \kappa_4 F_i IN4_i + u_i \qquad \textbf{(5.4)}$$

- The **population regression function for Model 5.4** is obtained by taking the conditional expectation of regression equation (5.4) for any given values of the regressors X_{i1}, X_{i2}, F_i, $IN2_i$, $IN3_i$, and $IN4_i$:

$$E(Y_i \mid X_{i1}, X_{i2}, F_i, IN2_i, IN3_i, IN4_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \lambda_f F_i + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i + \eta_1 F_i X_{i1} + \eta_2 F_i X_{i2} + \kappa_2 F_i IN2_i + \kappa_3 F_i IN3_i + \kappa_4 F_i IN4_i \qquad (5.4')$$

$$E(Y_i \mid X_{i1}, X_{i2}, F_i, IN2_i, IN3_i, IN4_i)$$

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \lambda_f F_i + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i$$

$$+ \eta_1 F_i X_{i1} + \eta_2 F_i X_{i2} + \kappa_2 F_i IN2_i + \kappa_3 F_i IN3_i + \kappa_4 F_i IN4_i \qquad (5.4')$$

- The ***female*** **population regression function for Model 5.4** is obtained by setting the female indicator $\mathbf{F_i = 1}$ **in (5.4')**:

$$E(Y_i \mid F_i = 1, X_{i1}, X_{i2}, IN2_i, IN3_i, IN4_i)$$

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \lambda_f + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i + \eta_1 X_{i1} + \eta_2 X_{i2} + \kappa_2 IN2_i + \kappa_3 IN3_i + \kappa_4 IN4_i$$

$$= (\beta_0 + \lambda_f) + (\beta_1 + \eta_1) X_{i1} + (\beta_2 + \eta_2) X_{i2} + (\pi_2 + \kappa_2) IN2_i + (\pi_3 + \kappa_3) IN3_i + (\pi_4 + \kappa_4) IN4_i \qquad \textbf{(5.4f)}$$

- The ***male*** **population regression function for Model 5.4** is obtained by setting the female indicator $\mathbf{F_i = 0}$ **in (5.4')**:

$$E(Y_i \mid F_i = 0, X_{i1}, X_{i2}, IN2_i, IN3_i, IN4_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i \qquad \textbf{(5.4m)}$$

- The ***female*** **population regression function for Model 5.4** is:

$$E(Y_i \mid F_i = 1, X_{i1}, X_{i2}, IN2_i, IN3_i, IN4_i)$$

$$= (\beta_0 + \lambda_f) + (\beta_1 + \eta_1)X_{i1} + (\beta_2 + \eta_2)X_{i2} + (\pi_2 + \kappa_2)IN2_i + (\pi_3 + \kappa_3)IN3_i + (\pi_4 + \kappa_4)IN4_i \quad \textbf{(5.4f)}$$

The ***female*** **population regression function for Model 5.4** implies that the **conditional mean value of Y for females *differs* across industries**:

1. The **conditional mean value of Y** for **females in *industry 1*** is:

$$E(Y_i \mid F_i = 1, X_{i1}, X_{i2}, IN2_i = 0, IN3 = 0_i, IN4_i = 0) = (\beta_0 + \lambda_f) + (\beta_1 + \eta_1)X_{i1} + (\beta_2 + \eta_2)X_{i2}$$

2. The **conditional mean value of Y** for **females in *industry 2*** is:

$$E(Y_i \mid F_i = 1, X_{i1}, X_{i2}, IN2_i = 1, IN3 = 0_i, IN4_i = 0) = (\beta_0 + \lambda_f) + (\beta_1 + \eta_1)X_{i1} + (\beta_2 + \eta_2)X_{i2} + (\pi_2 + \kappa_2)$$

3. The **conditional mean value of Y** for **females in *industry 3*** is:

$$E(Y_i \mid F_i = 1, X_{i1}, X_{i2}, IN2_i = 0, IN3 = 1_i, IN4_i = 0) = (\beta_0 + \lambda_f) + (\beta_1 + \eta_1)X_{i1} + (\beta_2 + \eta_2)X_{i2} + (\pi_3 + \kappa_3)$$

4. The **conditional mean value of Y** for **females in *industry 4*** is:

$$E(Y_i \mid F_i = 1, X_{i1}, X_{i2}, IN2_i = 0, IN3 = 0_i, IN4_i = 1) = (\beta_0 + \lambda_f) + (\beta_1 + \eta_1)X_{i1} + (\beta_2 + \eta_2)X_{i2} + (\pi_4 + \kappa_4)$$

- The ***male*** **population regression function for Model 5.4** is:

$$E(Y_i \mid F_i = 0, X_{i1}, X_{i2}, IN2_i, IN3_i, IN4_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i \qquad \textbf{(5.4m)}$$

The ***male*** **population regression function for Model 5.4** implies that the **conditional mean value of Y for males *differs* across industries**:

1. The **conditional mean value of Y** for **males in *industry 1*** is:

$$E(Y_i \mid F_i = 0, X_{i1}, X_{i2}, IN2_i = 0, IN3 = 0_i, IN4_i = 0) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2}$$

2. The **conditional mean value of Y** for **males in *industry 2*** is:

$$E(Y_i \mid F_i = 0, X_{i1}, X_{i2}, IN2_i = 1, IN3 = 0_i, IN4_i = 0) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \pi_2$$

3. The **conditional mean value of Y** for **males in *industry 3*** is:

$$E(Y_i \mid F_i = 0, X_{i1}, X_{i2}, IN2_i = 0, IN3 = 1_i, IN4_i = 0) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \pi_3$$

4. The **conditional mean value of Y** for **males in *industry 4*** is:

$$E(Y_i \mid F_i = 0, X_{i1}, X_{i2}, IN2_i = 0, IN3 = 0_i, IN4_i = 1) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \pi_4$$

- The ***female-male difference*** **in conditional mean Y for given values of the regressors** is obtained by **subtracting** the ***male*** **population regression function (5.4m)** from the ***female*** **population regression function (5.4f)**:

  The ***difference*** between the ***female*** **conditional mean Y** for *given* values of the regressors X_1, X_2, IN2, IN3, and IN4 and the ***male*** **conditional mean Y** for the *same* values of the regressors X_1, X_2, IN2, IN3, and IN4 is:

$$E(Y_i \mid F_i = 1, x_i^T) - E(Y_i \mid F_i = 0, x_i^T)$$

$$= \beta_0 + (\beta_1 + \eta_1)X_{i1} + (\beta_2 + \eta_2)X_{i2} + \lambda_f + (\pi_2 + \kappa_2)IN2_i + (\pi_3 + \kappa_3)IN3_i + (\pi_4 + \kappa_4)IN4_i$$
$$- (\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i)$$

$$= \beta_0 + (\beta_1 + \eta_1)X_{i1} + (\beta_2 + \eta_2)X_{i2} + \lambda_f + (\pi_2 + \kappa_2)IN2_i + (\pi_3 + \kappa_3)IN3_i + (\pi_4 + \kappa_4)IN4_i$$
$$- \beta_0 - \beta_1 X_{i1} - \beta_2 X_{i2} - \pi_2 IN2_i - \pi_3 IN3_i - \pi_4 IN4_i$$

$$= \beta_0 + \lambda_f + \beta_1 X_{i1} + \eta_1 X_{i1} + \beta_2 X_{i2} + \eta_2 X_{i2} + \pi_2 IN2_i + \kappa_2 IN2_i + \pi_3 IN3_i + \kappa_3 IN3_i + \pi_4 IN4_i + \kappa_4 IN4_i$$
$$- \beta_0 - \beta_1 X_{i1} - \beta_2 X_{i2} - \pi_2 IN2_i - \pi_3 IN3_i - \pi_4 IN4_i$$

$$= \lambda_f + \eta_1 X_{i1} + \eta_2 X_{i2} + \kappa_2 IN2_i + \kappa_3 IN3_i + \kappa_4 IN4_i \qquad \textbf{(5.4*)}$$

- Rewrite equation (5.4*) for the ***female-male difference*** in **conditional mean Y in Model 5.4** for given values of the regressors X_1, X_2, IN2, IN3, and IN4:

$$E(Y_i \mid F_i = 1, x_i^T) - E(Y_i \mid F_i = 0, x_i^T) = \lambda_f + \eta_1 X_{i1} + \eta_2 X_{i2} + \kappa_2 IN2_i + \kappa_3 IN3_i + \kappa_4 IN4_i \qquad \textbf{(5.4*)}$$

1. The ***female-male difference*** in **conditional mean Y for *industry 1*** for given values of X_1 and X_2 is obtained by setting $IN2_i = 0$ and $IN3_i = 0$ and $IN4_i = 0$ in (5.4*):

$$E(Y_i \mid F_i = 1, IN2_i = 0, IN3_i = 0, IN4_i = 0) - E(Y_i \mid F_i = 0, IN2_i = 0, IN3_i = 0, IN4_i = 0)$$

$$= \lambda_f + \eta_1 X_{i1} + \eta_2 X_{i2}$$

2. The ***female-male difference*** in **conditional mean Y for *industry 2*** for given values of X_1 and X_2 is obtained by setting $IN2_i = 1$ and $IN3_i = 0$ and $IN4_i = 0$ in (5.4*):

$$E(Y_i \mid F_i = 1, IN2_i = 1, IN3_i = 0, IN4_i = 0) - E(Y_i \mid F_i = 0, IN2_i = 1, IN3_i = 0, IN4_i = 0)$$

$$= \lambda_f + \eta_1 X_{i1} + \eta_2 X_{i2} + \kappa_2$$

3. The ***female-male difference*** in **conditional mean Y for *industry 3*** for given values of X_1 and X_2 is obtained by setting $IN2_i = 0$ and $IN3_i = 1$ and $IN4_i = 0$ in (5.4*):

$$E(Y_i \mid F_i = 1, IN2_i = 0, IN3_i = 1, IN4_i = 0) - E(Y_i \mid F_i = 0, IN2_i = 0, IN3_i = 1, IN4_i = 0)$$

$$= \lambda_f + \eta_1 X_{i1} + \eta_2 X_{i2} + \kappa_3$$

$$E(Y_i \mid F_i = 1, x_i^T) - E(Y_i \mid F_i = 0, x_i^T) = \lambda_f + \eta_1 X_{i1} + \eta_2 X_{i2} + \kappa_2 IN2_i + \kappa_3 IN3_i + \kappa_4 IN4_i \qquad \textbf{(5.4*)}$$

4. The ***female-male difference*** in **conditional mean Y for *industry 4*** for given values of X_1 and X_2 is obtained by setting $IN2_i = 0$ and $IN3_i = 0$ and $IN4_i = 1$ in (5.4*):

$$E(Y_i \mid F_i = 1, IN2_i = 0, IN3_i = 0, IN4_i = 1) - E(Y_i \mid F_i = 0, IN2_i = 0, IN3_i = 0, IN4_i = 1)$$

$$= \lambda_f + \eta_1 X_{i1} + \eta_2 X_{i2} + \kappa_4$$

Propositions to Test in Model 5.4 Respecting Female-Male Differences in Conditional Mean Y

- The ***female-male difference*** in **conditional mean Y** in Model 5.4 for given values of the regressors X_1, X_2, IN2, IN3, and IN4:

$$E(Y_i \mid F_i = 1, x_i^T) - E(Y_i \mid F_i = 0, x_i^T) = \lambda_f + \eta_1 X_{i1} + \eta_2 X_{i2} + \kappa_2 IN2_i + \kappa_3 IN3_i + \kappa_4 IN4_i \qquad \textbf{(5.4*)}$$

<u>Test 1</u>: The ***female-male difference* in conditional mean Y *equals zero* for all observations**.

H$_0$: $\lambda_f = 0$ *and* $\eta_1 = 0$ *and* $\eta_2 = 0$ *and* $\kappa_2 = 0$ *and* $\kappa_3 = 0$ *and* $\kappa_4 = 0$

H$_1$: $\lambda_f \neq 0$ *and/or* $\eta_1 \neq 0$ *and/or* $\eta_2 \neq 0$ *and/or* $\kappa_2 \neq 0$ *and/or* $\kappa_3 \neq 0$ *and/or* $\kappa_4 \neq 0$

<u>Test 2</u>: The ***female-male difference* in conditional mean Y *equals a constant, i.e.,* it does not depend on *industry* or on *the values of X_1 and X_2*.**

H$_0$: $\eta_1 = 0$ *and* $\eta_2 = 0$ *and* $\kappa_2 = 0$ *and* $\kappa_3 = 0$ *and* $\kappa_4 = 0$

H$_1$: $\eta_1 \neq 0$ *and/or* $\eta_2 \neq 0$ *and/or* $\kappa_2 \neq 0$ *and/or* $\kappa_3 \neq 0$ *and/or* $\kappa_4 \neq 0$

- The ***female-male difference*** in **conditional mean Y** in Model 5.4 for given values of the regressors X_1, X_2, IN2, IN3, and IN4:

$$E(Y_i \mid F_i = 1, x_i^T) - E(Y_i \mid F_i = 0, x_i^T) = \lambda_f + \eta_1 X_{i1} + \eta_2 X_{i2} + \kappa_2 IN2_i + \kappa_3 IN3_i + \kappa_4 IN4_i \qquad \textbf{(5.4*)}$$

Test 3: The ***female-male difference*** **in conditional mean Y does not depend on *industry*** – i.e., industry effects are identical for females and males.

H_0: $\kappa_2 = 0$ *and* $\kappa_3 = 0$ *and* $\kappa_4 = 0$

H_1: $\kappa_2 \neq 0$ *and/or* $\kappa_3 \neq 0$ *and/or* $\kappa_4 \neq 0$

Test 4: The ***female-male difference*** **in conditional mean Y does not depend on *the values of the regressors X_1 and X_2*.**

H_0: $\eta_1 = 0$ *and* $\eta_2 = 0$

H_1: $\eta_1 \neq 0$ *and/or* $\eta_2 \neq 0$

Tests for Industry Effects in Model 5.4

Test 1: Test for Industry Effects for *Females* in Model 5.4

- The **population regression function for Model 5.4** gives the conditional mean value of Y for any given values of the regressors X_{i1}, X_{i2}, F_i, $IN2_i$, $IN3_i$, and $IN4_i$:

$$E(Y_i \mid X_{i1}, X_{i2}, F_i, IN2_i, IN3_i, IN4_i)$$

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \lambda_f F_i + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i$$
$$+ \eta_1 F_i X_{i1} + \eta_2 F_i X_{i2} + \kappa_2 F_i IN2_i + \kappa_3 F_i IN3_i + \kappa_4 F_i IN4_i \qquad (5.4')$$

- The ***female*** **population regression function for Model 5.4** is obtained by setting the female indicator **$F_i = 1$ in (5.4')**:

$$E(Y_i \mid F_i = 1, X_{i1}, X_{i2}, IN2_i, IN3_i, IN4_i)$$

$$= \beta_0 + \lambda_f + (\beta_1 + \eta_1)X_{i1} + (\beta_2 + \eta_2)X_{i2} + (\pi_2 + \kappa_2)IN2_i + (\pi_3 + \kappa_3)IN3_i + (\pi_4 + \kappa_4)IN4_i \qquad \textbf{… (5.4f)}$$

Hypothesis Test: Test the proposition of **no industry effects for *females*** – i.e., the proposition that there are no inter-industry differences in conditional mean Y values for females.

H_0: $\pi_2 + \kappa_2 = 0$ *and* $\pi_3 + \kappa_3 = 0$ *and* $\pi_4 + \kappa_4 = 0$

H_1: $\pi_2 + \kappa_2 \neq 0$ *and/or* $\pi_3 + \kappa_3 \neq 0$ *and/or* $\pi_4 + \kappa_4 \neq 0$

Restricted Model for Females: Substitute the three restrictions specified by H_0 into the unrestricted female regression function (5.4f) to get the ***restricted* female regression function with *no* industry effects**.

$$E(Y_i \mid F_i = 1, X_{i1}, X_{i2}, IN2_i, IN3_i, IN4_i)$$

$$= \beta_0 + \lambda_f + (\beta_1 + \eta_1)X_{i1} + (\beta_2 + \eta_2)X_{i2} + (\pi_2 + \kappa_2)IN2_i + (\pi_3 + \kappa_3)IN3_i + (\pi_4 + \kappa_4)IN4_i \qquad \textbf{... (5.4f)}$$

Setting $\pi_2 + \kappa_2 = 0$ and $\pi_3 + \kappa_3 = 0$ and $\pi_4 + \kappa_4 = 0$ in (5.4f) yields the ***restricted* female regression function with *no* industry effects**:

$$E(Y_i \mid F_i = 1, X_{i1}, X_{i2}, IN2_i, IN3_i, IN4_i) = \beta_0 + \lambda_f + (\beta_1 + \eta_1)X_{i1} + (\beta_2 + \eta_2)X_{i2}$$

<u>Test 2</u>: Test for Industry Effects for *Males* in Model 5.4

- The ***male* population regression function for Model 5.4** is obtained by setting the female indicator $\mathbf{F_i = 0}$ **in (5.4')**:

$$E(Y_i \mid X_{i1}, X_{i2}, F_i, IN2_i, IN3_i, IN4_i)$$

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \lambda_f F_i + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i$$
$$+ \eta_1 F_i X_{i1} + \eta_2 F_i X_{i2} + \kappa_2 F_i IN2_i + \kappa_3 F_i IN3_i + \kappa_4 F_i IN4_i \qquad (5.4')$$

$$E(Y_i \mid F_i = 0, X_{i1}, X_{i2}, IN2_i, IN3_i, IN4_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i \qquad \mathbf{(5.4m)}$$

Hypothesis Test: Test the proposition of **no industry effects for *males*** – i.e., the proposition that there are no inter-industry differences in conditional mean Y values for males.

H_0: $\pi_2 = 0$ *and* $\pi_3 = 0$ *and* $\pi_4 = 0$

H_1: $\pi_2 \neq 0$ *and/or* $\pi_3 \neq 0$ *and/or* $\pi_4 \neq 0$

Restricted Model for Males: Substitute the three restrictions specified by H_0 into the unrestricted male regression function (5.4f) to get the ***restricted* male regression function with *no* industry effects**.

$$E(Y_i \mid F_i = 0, X_{i1}, X_{i2}, IN2_i, IN3_i, IN4_i)$$

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} \qquad \mathbf{\ldots\ (5.3m)}$$

<u>Test 3</u>: Test for *Female-Male Differences* in Industry Effects in Model 5.4

Proposition: Test the proposition of ***equal* industry effects for *females* and *males*** – i.e., the proposition that the inter-industry differences in conditional mean Y values for females are identical to the inter-industry differences in conditional mean Y values for males.

Null and Alternative Hypotheses:

H_0: $\kappa_2 = 0$ *and* $\kappa_3 = 0$ *and* $\kappa_4 = 0$

H_1: $\kappa_2 \neq 0$ *and/or* $\kappa_3 \neq 0$ *and/or* $\kappa_4 \neq 0$

Restricted Model: Substitute the three restrictions specified by H_0 into the unrestricted pooled regression function (5.4') for Model 5.4.

$$E(Y_i \mid X_{i1}, X_{i2}, F_i, IN2_i, IN3_i, IN4_i)$$

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \lambda_f F_i + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i + \eta_1 F_i X_{i1} + \eta_2 F_i X_{i2} + \kappa_2 F_i IN2_i + \kappa_3 F_i IN3_i + \kappa_4 F_i IN4_i \qquad (5.4')$$

The ***restricted* pooled regression function** under H_0 is therefore:

$$E(Y_i \mid X_{i1}, X_{i2}, F_i, IN2_i, IN3_i, IN4_i)$$

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \lambda_f F_i + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i + \eta_1 F_i X_{i1} + \eta_2 F_i X_{i2}$$

- ***Interpretation of the coefficients in Model 5.4*** for which the PRF (population regression function) is:

$$E(Y_i \mid X_{i1}, X_{i2}, F_i, IN2_i, IN3_i, IN4_i)$$

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \lambda_f F_i + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i$$

$$+ \eta_1 F_i X_{i1} + \eta_2 F_i X_{i2} + \kappa_2 F_i IN2_i + \kappa_3 F_i IN3_i + \kappa_4 F_i IN4_i \qquad \textbf{(5.4')}$$

β_0 = intercept for ***males*** **in industry 1**

$\beta_0 + \lambda_f$ = intercept for ***females*** **in industry 1**

λ_f = ***female*** **industry 1** intercept – ***male*** **industry 1** intercept

$\beta_0 + \pi_2$ = intercept for ***males*** **in industry 2**

$\beta_0 + \lambda_f + \pi_2 + \kappa_2$ = intercept for ***females*** **in industry 2**

$\lambda_f + \kappa_2$ = ***female*** **industry 2** intercept – ***male*** **industry 2** intercept

$\beta_0 + \pi_3$ = intercept for ***males*** **in industry 3**

$\beta_0 + \lambda_f + \pi_3 + \kappa_3$ = intercept for ***females*** **in industry 3**

$\lambda_f + \kappa_3$ = ***female*** **industry 3** intercept – ***male*** **industry 3** intercept

$\beta_0 + \pi_4$ = intercept for ***males*** **in industry 4**

$\beta_0 + \lambda_f + \pi_4 + \kappa_4$ = intercept for ***females*** **in industry 4**

$\lambda_f + \kappa_4$ = ***female*** **industry 4** intercept – ***male*** **industry 4** intercept

- ***Interpretation of the coefficients in Model 5.4*** for which the PRF (population regression function) is:

$$E(Y_i \mid X_{i1}, X_{i2}, F_i, IN2_i, IN3_i, IN4_i)$$

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \lambda_f F_i + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i$$

$$+ \eta_1 F_i X_{i1} + \eta_2 F_i X_{i2} + \kappa_2 F_i IN2_i + \kappa_3 F_i IN3_i + \kappa_4 F_i IN4_i \qquad \textbf{(5.4')}$$

β_1 = the **marginal effect of X_1** for ***males***
$\beta_1 + \eta_1$ = the **marginal effect of X_1** for ***females***
η_1 = the ***female-male difference*** in the **marginal effect of X_1**

β_2 = the **marginal effect of X_2** for ***males***
$\beta_2 + \eta_2$ = the **marginal effect of X_2** for ***females***
η_2 = the ***female-male difference*** in the **marginal effect of X_2**

β_3 = the **marginal effect of X_3** for ***males***
$\beta_3 + \eta_3$ = the **marginal effect of X_3** for ***females***
η_3 = the ***female-male difference*** in the **marginal effect of X_3**

- ***Interpretation of the coefficients κ_2, κ_3 and κ_4 in Model 5.4***, for which the PRF is

$$E(Y_i \mid X_{i1}, X_{i2}, F_i, IN2_i, IN3_i, IN4_i)$$
$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \lambda_f F_i + \pi_2 IN2_i + \pi_3 IN3_i + \pi_4 IN4_i$$
$$+ \eta_1 F_i X_{i1} + \eta_2 F_i X_{i2} + \kappa_2 F_i IN2_i + \kappa_3 F_i IN3_i + \kappa_4 F_i IN4_i \qquad \textbf{(5.4')}$$

Two alternative interpretations of the coefficient κ_2 in Model 5.4

Both use the **conditional mean values of Y** for **four subgroups**:

1. The **conditional mean value of Y** for **females in *industry 2***: set $F_i = 1$, $IN2_i = 1$, $IN3_i = 0$, $IN4_i = 0$ in (5.4')

$$E(Y_i \mid F_i = 1, X_{i1}, X_{i2}, IN2_i = 1, IN3 = 0_i, IN4_i = 0) = (\beta_0 + \lambda_f) + (\beta_1 + \eta_1)X_{i1} + (\beta_2 + \eta_2)X_{i2} + (\pi_2 + \kappa_2)$$

2. The **conditional mean value of Y** for ***males* in *industry 2***: set $F_i = 0$, $IN2_i = 1$, $IN3_i = 0$, $IN4_i = 0$ in (5.4')

$$E(Y_i \mid F_i = 0, X_{i1}, X_{i2}, IN2_i = 1, IN3 = 0_i, IN4_i = 0) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \pi_2$$

3. The **conditional mean value of Y** for **females in *industry 1***: set $F_i = 1$, $IN2_i = 0$, $IN3_i = 0$, $IN4_i = 0$ in (5.4')

$$E(Y_i \mid X_{i1}, X_{i2}, F_i = 1, IN2_i = 0, IN3_i = 0, IN4_i = 0) = (\beta_0 + \lambda_f) + (\beta_1 + \eta_1)X_{i1} + (\beta_2 + \eta_2)X_{i2}$$

4. The **conditional mean value of Y** for ***males* in *industry 1***: set $F_i = 0$, $IN2_i = 0$, $IN3_i = 0$, $IN4_i = 0$ in (5.4'),

$$E(Y_i \mid X_{i1}, X_{i2}, F_i = 0, IN2_i = 0, IN3_i = 0, IN4_i = 0) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2}$$

<u>Interpretation 1</u> of κ_2 in Model 5.4

(1) Female-male difference in conditional mean Y **for *industry 2***:

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \lambda_f + \pi_2 + \eta_1 X_{i1} + \eta_2 X_{i2} + \kappa_2$$

$$- \beta_0 - \beta_1 X_{i1} - \beta_2 X_{i2} - \pi_2$$

$$= \lambda_f + \eta_1 X_{i1} + \eta_2 X_{i2} + \kappa_2$$

(2) Female-male difference in conditional mean Y **for *industry 1***:

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \lambda_f + \eta_1 X_{i1} + \eta_2 X_{i2}$$

$$- \beta_0 - \beta_1 X_{i1} - \beta_2 X_{i2}$$

$$= \lambda_f + \eta_1 X_{i1} + \eta_2 X_{i2}$$

Subtract (2) from (1):

Difference-in-differences* = female-male difference** in conditional mean Y **for *industry 2* <u>*minus*</u> female-male difference** in conditional mean Y **for *industry 1:

$$= \lambda_f + \eta_1 X_{i1} + \eta_2 X_{i2} + \kappa_2 - \lambda_f - \eta_1 X_{i1} - \eta_2 X_{i2} = \kappa_2$$

<u>Interpretation 2</u> of κ_2 in Model 5.4

(1) Industry 2-industry 1 difference in conditional mean Y for ***females***:

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \lambda_f + \pi_2 + \eta_1 X_{i1} + \eta_2 X_{i2} + \kappa_2$$
$$- \beta_0 - \beta_1 X_{i1} - \beta_2 X_{i2} - \lambda_f - \eta_1 X_{i1} - \eta_2 X_{i2}$$

$$= \pi_2 + \kappa_2$$

(2) Industry 2-industry 1 difference in conditional mean Y for ***males***:

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \pi_2 - \beta_0 - \beta_1 X_{i1} - \beta_2 X_{i2} = \pi_2$$

Subtract (2) from (1):

Difference-in-differences* = industry 2-industry 1 difference** in conditional mean Y **for *females* <u>*minus*</u> industry 2-industry 1 difference** in conditional mean Y **for *males:

$$= \pi_2 + \kappa_2 - \pi_2 = \kappa_2$$

Standard Notation for *Model 5.4*

We now reformulate Model 5.4, developed in the previous section, in much cleaner and more conventional notation.

- The **population regression equation for Model 5.4** can be written in more standard notation as:

$$\begin{aligned} Y_i = {} & \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 IN2_i + \beta_4 IN3_i + \beta_5 IN4_i + \delta_0 F_i \\ & + \delta_1 F_i X_{i1} + \delta_2 F_i X_{i2} + \delta_3 F_i IN2_i + \delta_4 F_i IN3_i + \delta_5 F_i IN4_i + u_i \end{aligned} \qquad \textbf{(5.4)}$$

- The **population regression *function* for Model 5.4** is:

$$\begin{aligned} & E(Y_i \,|\, X_{i1}, X_{i2}, F_i, IN2_i, IN3_i, IN4_i) \\ & = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 IN2_i + \beta_4 IN3_i + \beta_5 IN4_i + \delta_0 F_i \\ & \quad + \delta_1 F_i X_{i1} + \delta_2 F_i X_{i2} + \delta_3 F_i IN2_i + \delta_4 F_i IN3_i + \delta_5 F_i IN4_i \end{aligned} \qquad (5.4')$$

$$E(Y_i \mid X_{i1}, X_{i2}, F_i, IN2_i, IN3_i, IN4_i)$$
$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 IN2_i + \beta_4 IN3_i + \beta_5 IN4_i + \delta_0 F_i$$
$$+ \delta_1 F_i X_{i1} + \delta_2 F_i X_{i2} + \delta_3 F_i IN2_i + \delta_4 F_i IN3_i + \delta_5 F_i IN4_i \quad (5.4')$$

- The ***female*** **population regression function for Model 5.4** is obtained by setting the female indicator $F_i = 1$ in (5.4'):

$$E(Y_i \mid F_i = 1, X_{i1}, X_{i2}, IN2_i, IN3_i, IN4_i)$$
$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 IN2_i + \beta_4 IN3_i + \beta_5 IN4_i$$
$$+ \delta_0 + \delta_1 X_{i1} + \delta_2 X_{i2} + \delta_3 IN2_i + \delta_4 IN3_i + \delta_5 IN4_i \quad \textbf{(5.4f)}$$
$$= (\beta_0 + \delta_0) + (\beta_1 + \delta_1) X_{i1} + (\beta_2 + \delta_2) X_{i2} + (\beta_3 + \delta_3) IN2_i + (\beta_4 + \delta_4) IN3_i + (\beta_5 + \delta_5) IN4_i \quad \textbf{(5.4f)}$$

- The ***male*** **population regression function for Model 5.4** is obtained by setting the female indicator $F_i = 0$ in (5.4'):

$$E(Y_i \mid F_i = 0, X_{i1}, X_{i2}, IN2_i, IN3_i, IN4_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 IN2_i + \beta_4 IN3_i + \beta_5 IN4_i \quad \textbf{(5.4m)}$$

- The ***female-male difference* in conditional mean Y for Model 5.4** is obtained by subtracting the male regression function (5.4m) from the female regression function (5.4f):

The ***female* population regression function for Model 5.4** is:

$$E(Y_i \mid F_i = 1, X_{i1}, X_{i2}, IN2_i, IN3_i, IN4_i)$$

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 IN2_i + \beta_4 IN3_i + \beta_5 IN4_i$$
$$+ \delta_0 + \delta_1 X_{i1} + \delta_2 X_{i2} + \delta_3 IN2_i + \delta_4 IN3_i + \delta_5 IN4_i \qquad \textbf{(5.4f)}$$

$$= (\beta_0 + \delta_0) + (\beta_1 + \delta_1) X_{i1} + (\beta_2 + \delta_2) X_{i2} + (\beta_3 + \delta_3) IN2_i + (\beta_4 + \delta_4) IN3_i + (\beta_5 + \delta_5) IN4_i \qquad \textbf{(5.4f)}$$

The ***male* population regression function for Model 5.4** is:

$$E(Y_i \mid F_i = 0, X_{i1}, X_{i2}, IN2_i, IN3_i, IN4_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 IN2_i + \beta_4 IN3_i + \beta_5 IN4_i \qquad \textbf{(5.4m)}$$

The ***female-male difference* in conditional mean Y for Model 5.4** is therefore:

$$E(Y_i \mid F_i = 1, x_i^T) - E(Y_i \mid F_i = 0, x_i^T)$$

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 IN2_i + \beta_4 IN3_i + \beta_5 IN4_i$$
$$+ \delta_0 + \delta_1 X_{i1} + \delta_2 X_{i2} + \delta_3 IN2_i + \delta_4 IN3_i + \delta_5 IN4_i$$
$$- \beta_0 - \beta_1 X_{i1} - \beta_2 X_{i2} - \beta_3 IN2_i - \beta_4 IN3_i - \beta_5 IN4_i$$

$$= \delta_0 + \delta_1 X_{i1} + \delta_2 X_{i2} + \delta_3 IN2_i + \delta_4 IN3_i + \delta_5 IN4_i$$

- **Interpretation of regression coefficients** in the pooled **regression function for Model 5.4** given by (5.4'):

$$E(Y_i \mid X_{i1}, X_{i2}, F_i, IN2_i, IN3_i, IN4_i)$$
$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 IN2_i + \beta_4 IN3_i + \beta_5 IN4_i + \delta_0 F_i$$
$$+ \delta_1 F_i X_{i1} + \delta_2 F_i X_{i2} + \delta_3 F_i IN2_i + \delta_4 F_i IN3_i + \delta_5 F_i IN4_i \qquad (5.4')$$

The ***female*** **population regression function for Model 5.4** is:

$$E(Y_i \mid F_i = 1, X_{i1}, X_{i2}, IN2_i, IN3_i, IN4_i)$$
$$= (\beta_0 + \delta_0) + (\beta_1 + \delta_1)X_{i1} + (\beta_2 + \delta_2)X_{i2} + (\beta_3 + \delta_3)IN2_i + (\beta_4 + \delta_4)IN3_i + (\beta_5 + \delta_5)IN4_i \qquad \textbf{(5.4f)}$$

$\beta_j + \delta_j$ = the ***female*** **regression coefficient** on regressor j (j = 0, 1, …, 5)

The ***male*** **population regression function for Model 5.4** is:

$$E(Y_i \mid F_i = 0, X_{i1}, X_{i2}, IN2_i, IN3_i, IN4_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 IN2_i + \beta_4 IN3_i + \beta_5 IN4_i \qquad \textbf{(5.4m)}$$

β_j = the ***male*** **regression coefficient** on regressor j (j = 0, 1, …, 5)

The ***female-male difference*** **in conditional mean Y for Model 5.4** is:

$$E(Y_i \mid F_i = 1, x_i^T) - E(Y_i \mid F_i = 0, x_i^T) = \delta_0 + \delta_1 X_{i1} + \delta_2 X_{i2} + \delta_3 IN2_i + \delta_4 IN3_i + \delta_5 IN4_i$$

δ_j = the ***female*** **regression coefficient** on regressor j

minus

the ***male*** **regression coefficient** on regressor j

= the ***female-male*** **coefficient** ***difference*** for regressor j (j = 0, 1, …, 6)

- ***Stata* commands for computing the female coefficient estimates in Model 5.4**

The ***female* OLS sample regression function for Model 5.4** is:

$$\hat{E}(Y_i \mid F_i = 1, X_{i1}, X_{i2}, IN2_i, IN3_i, IN4_i)$$

$$= (\hat{\beta}_0 + \hat{\delta}_0) + (\hat{\beta}_1 + \hat{\delta}_1)X_{i1} + (\hat{\beta}_2 + \hat{\delta}_2)X_{i2} + (\hat{\beta}_3 + \hat{\delta}_3)IN2_i + (\hat{\beta}_4 + \hat{\delta}_4)IN3_i + (\hat{\beta}_5 + \hat{\delta}_5)IN4_i$$

The following *Stata* commands compute the OLS female coefficient estimates for Model 5.4:

`lincom _b[_cons] + _b[f]` computes $\hat{\beta}_0 + \hat{\delta}_0$

`lincom _b[x1] + _b[fx1]` computes $\hat{\beta}_1 + \hat{\delta}_1$

`lincom _b[x2] + _b[fx2]` computes $\hat{\beta}_2 + \hat{\delta}_2$

`lincom _b[in2] + _b[fin2]` computes $\hat{\beta}_3 + \hat{\delta}_3$

`lincom _b[in3] + _b[fin3]` computes $\hat{\beta}_4 + \hat{\delta}_4$

`lincom _b[in4] + _b[fin4]` computes $\hat{\beta}_5 + \hat{\delta}_5$

- The ***female*** **population regression function for Model 5.4** gives the conditional mean value of Y for females with given values of the regressors:

$$E(Y_i \mid F_i = 1, X_{i1}, X_{i2}, IN2_i, IN3_i, IN4_i)$$

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 IN2_i + \beta_4 IN3_i + \beta_5 IN4_i \; + \delta_0 + \delta_1 X_{i1} + \delta_2 X_{i2} + \delta_3 IN2_i + \delta_4 IN3_i + \delta_5 IN4_i \qquad \textbf{(5.4f)}$$

$$= (\beta_0 + \delta_0) + (\beta_1 + \delta_1) X_{i1} + (\beta_2 + \delta_2) X_{i2} + (\beta_3 + \delta_3) IN2_i + (\beta_4 + \delta_4) IN3_i + (\beta_5 + \delta_5) IN4_i \qquad \textbf{(5.4f)}$$

(1) Conditional mean Y for ***females*** **in** ***industry 1*** is:

$$E(Y_i \mid F_i = 1, X_{i1}, X_{i2}, IN2_i = 0, IN3_i = 0, IN4_i = 0)$$

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \delta_0 + \delta_1 X_{i1} + \delta_2 X_{i2}$$

$$= \beta_0 + \delta_0 + (\beta_1 + \delta_1) X_{i1} + (\beta_2 + \delta_2) X_{i2}$$

(2) Conditional mean Y for ***females*** **in** ***industry 2*** is:

$$E(Y_i \mid F_i = 1, X_{i1}, X_{i2}, IN2_i = 1, IN3_i = 0, IN4_i = 0)$$

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 + \delta_0 + \delta_1 X_{i1} + \delta_2 X_{i2} + \delta_3$$

$$= \beta_0 + \delta_0 + (\beta_1 + \delta_1) X_{i1} + (\beta_2 + \delta_2) X_{i2} + (\beta_3 + \delta_3)$$

$$E(Y_i \mid F_i = 1, X_{i1}, X_{i2}, IN2_i, IN3_i, IN4_i)$$

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 IN2_i + \beta_4 IN3_i + \beta_5 IN4_i + \delta_0 + \delta_1 X_{i1} + \delta_2 X_{i2} + \delta_3 IN2_i + \delta_4 IN3_i + \delta_5 IN4_i \quad \textbf{(5.4f)}$$

$$= (\beta_0 + \delta_0) + (\beta_1 + \delta_1) X_{i1} + (\beta_2 + \delta_2) X_{i2} + (\beta_3 + \delta_3) IN2_i + (\beta_4 + \delta_4) IN3_i + (\beta_5 + \delta_5) IN4_i \quad \textbf{(5.4f)}$$

(3) Conditional mean Y for ***females*** **in** ***industry 3*** is:

$$E(Y_i \mid F_i = 1, X_{i1}, X_{i2}, IN2_i = 0, IN3_i = 1, IN4_i = 0)$$

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_4 + \delta_0 + \delta_1 X_{i1} + \delta_2 X_{i2} + \delta_4$$

$$= \beta_0 + \delta_0 + (\beta_1 + \delta_1) X_{i1} + (\beta_2 + \delta_2) X_{i2} + (\beta_4 + \delta_4)$$

(4) Conditional mean Y for ***females*** **in** ***industry 4*** is:

$$E(Y_i \mid F_i = 1, X_{i1}, X_{i2}, IN2_i = 0, IN3_i = 0, IN4_i = 1)$$

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_5 + \delta_0 + \delta_1 X_{i1} + \delta_2 X_{i2} + \delta_5$$

$$= \beta_0 + \delta_0 + (\beta_1 + \delta_1) X_{i1} + (\beta_2 + \delta_2) X_{i2} + (\beta_5 + \delta_5)$$

- The ***male*** **population regression function for Model 5.4** gives the conditional mean value of Y for males with given values of the regressors:

$$E(Y_i \mid F_i = 0, X_{i1}, X_{i2}, IN2_i, IN3_i, IN4_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 IN2_i + \beta_4 IN3_i + \beta_5 IN4_i \quad \textbf{(5.4m)}$$

(1) Conditional mean Y for ***males*** **in** ***industry 1*** is:

$$E(Y_i \mid F_i = 0, X_{i1}, X_{i2}, IN2_i = 0, IN3_i = 0, IN4_i = 0) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2}$$

(2) Conditional mean Y for ***males*** **in** ***industry 2*** is:

$$E(Y_i \mid F_i = 0, X_{i1}, X_{i2}, IN2_i = 1, IN3_i = 0, IN4_i = 0) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3$$

$$= (\beta_0 + \beta_3) + \beta_1 X_{i1} + \beta_2 X_{i2}$$

(3) Conditional mean Y for ***males*** **in** ***industry 3*** is:

$$E(Y_i \mid F_i = 0, X_{i1}, X_{i2}, IN2_i = 0, IN3_i = 1, IN4_i = 0) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_4$$

$$= (\beta_0 + \beta_4) + \beta_1 X_{i1} + \beta_2 X_{i2}$$

(4) Conditional mean Y for ***males*** **in** ***industry 4*** is:

$$E(Y_i \mid F_i = 0, X_{i1}, X_{i2}, IN2_i = 0, IN3_i = 0, IN4_i = 1) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_5$$

$$= (\beta_0 + \beta_5) + \beta_1 X_{i1} + \beta_2 X_{i2}$$