# ECON 452* -- NOTE 5

## Using Dummy Variable Regressors for Two-Category Categorical Variables

### ❐ Nature and Properties of Indicator (Dummy) Variables

- **Indicator (or dummy) variables** are ***binary*** **variables** -- i.e., **variables that take *only two values***.

  The value **1** indicates **the presence** of some characteristic or attribute.

  The value **0** indicates **the absence** of that same characteristic or attribute.

- **Consider** a **two-way partitioning** of a population or sample into **two *mutually exclusive* and *exhaustive* subsets or groups** -- ***females*** and ***males***.

  - Let $\mathbf{F_i}$ be the **female indicator (dummy) variable**, defined as follows:

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

  - Let $\mathbf{M_i}$ be the **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 Indicator Variables $F_i$ and $M_i$**

For each and every i (population member or sample observation):

if $F_i = 1$ then $M_i = 0$

and

if $M_i = 1$ then $F_i = 0$.

The definition of the indicator variables $F_i$ and $M_i$ thus implies that they satisfy the following **adding-up property**:

$$\mathbf{F_i + M_i = 1 \qquad \forall\ i.}$$

- **Implications of the Adding-Up Property**

1. **Only *one* of the two dummy variables $F_i$ and $M_i$** is required to ***completely represent* the *two-way partitioning*** of a population and sample into females and males.

    - given $M_i$ values, the adding-up property implies that $F_i = 1 - M_i$.
    - given $F_i$ values, the adding-up property implies that $M_i = 1 - F_i$.

**2. <u>General Rule</u>:** A ***categorical* variable with n categories** can be completely represented by **a set of n–1 indicator (dummy) variables**.

The **general adding-up property** states that

$$D1_i + D2_i + D3_i + \cdots + Dn_i = 1 \quad \forall \; i.$$

- ***Example:*** Consider a categorical variable $INDUSTRY_i$ representing individual employees' **industry sector** of employment. $INDUSTRY_i$ is defined as follows:

$INDUSTRY_i$ = 1 if person i is employed in construction industries;
= 2 if person i is employed in nondurable manufacturing industries;
= 3 if person i is employed in durable manufacturing industries;
= 4 if person i is employed in transportation, communications, or public utilities industries;
= 5 if person i is employed in wholesale or retail trades;
= 6 if person i is employed in services industries;
= 7 if person i is employed in professional services industries.

- Define a set of industry sector dummy variables to represent the categorical variable $\text{INDUSTRY}_i$.

  $\text{construc}_i$ = 1 if person i is employed in construction industries, = 0 otherwise;
  $\text{ndurman}_i$ = 1 if person i is employed in nondurable manufacturing, = 0 otherwise;
  $\text{durman}_i$ = 1 if person i is employed in durable manufacturing, = 0 otherwise;
  $\text{trcommpu}_i$ = 1 if person i is employed in transportation, communications, or public utilities, = 0 otherwise;
  $\text{trade}_i$ = 1 if person i is employed in wholesale or retail trades, = 0 otherwise;
  $\text{services}_i$ = 1 if person i is employed in services industries, = 0 otherwise;
  $\text{profserv}_i$ = 1 if person i is employed in professional services, = 0 otherwise.

- By definition, the seven industry sector dummy variables satisfy the ***adding-up property***:

  $$\text{construc}_i + \text{ndurman}_i + \text{durman}_i + \text{trcommpu}_i + \text{trade}_i + \text{services}_i + \text{profserv}_i = 1 \qquad \forall\ i.$$

- **<u>Implication of the adding-up property</u>**: The partitioning of the population or sample into ***seven*** **mutually exclusive and exhaustive industry sector groups** can be completely represented by ***any six*** **of the seven industry sector dummy variables** $\boldsymbol{construc_i}$, $\boldsymbol{ndurman_i}$, $\boldsymbol{durman_i}$, $\boldsymbol{trcommpu_i}$, $\boldsymbol{trade_i}$, $\boldsymbol{services_i}$, and $\boldsymbol{profserv_i}$.

  For example, the industry dummy variable $\text{durman}_i$ can be computed from the other six industry sector dummy variables as follows:

  $$\text{durman}_i = 1 - \text{construc}_i - \text{ndurman}_i - \text{trcommpu}_i - \text{trade}_i - \text{services}_i - \text{profserv}_i \qquad \forall\ i.$$

  If durable manufacturing industries are chosen as the ***base group***, or ***reference group***, for the categorical variable ***industry***, then the durable manufacturing dummy variable $\boldsymbol{durman_i}$ would be excluded from the set of dummy variable regressors used to represent ***industry*** in a linear regression equation.

## ❒ Indicator Variables as Additive Regressors: Differences in Intercepts

***Nature:*** When indicator (dummy) variables are introduced additively as additional regressors in linear regression models, they allow for **different *intercept* coefficients** across identifiable subsets of observations in the population.

***Example:*** Suppose we have two mutually exclusive and exhaustive subgroups of observations in the relevant population -- ***females*** **and** ***males***.

We distinguish between these two subgroups of observations by using a ***female*** **indicator variable** $\mathbf{F_i}$ defined as follows:

$$F_i = 1 \text{ if observation i is female}$$
$$\quad = 0 \text{ if observation i is not female (i.e., is male).}$$

***Model 1:*** Contains five regressors in the two explanatory variables $X_1$ and $X_2$, both of which are *continuous*.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}^2 + \beta_4 X_{i2}^2 + \beta_5 X_{i1} X_{i2} + u_i \qquad \textbf{(1)}$$

- The population regression function, or conditional mean function, $f(X_{i1}, X_{i2})$ in Model 1 takes the form

$$E(Y_i \mid X_{i1}, X_{i2}) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}^2 + \beta_4 X_{i2}^2 + \beta_5 X_{i1} X_{i2}.$$

- Model 1 **does not allow for any coefficient differences** between males and females.
  - ♦ Model 1 assumes that all six regression coefficients $\beta_j$ (j = 0, 1, …, 5) are the same for males and females.
  - ♦ Model 1 assumes that the population regression function is the same for both females and males.

***Model 2:*** Allows for **different male and female *intercepts*** by introducing the female indicator variable $F_i$ as an additional additive regressor in Model 1.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}^2 + \beta_4 X_{i2}^2 + \beta_5 X_{i1} X_{i2} + \delta_0 F_i + u_i \quad \textbf{(2)}$$

- The **population regression function**, or conditional mean function, **for Model 2** is obtained by taking the conditional expectation of regression equation (2) for any given values of the three explanatory variables $X_{i1}$, $X_{i2}$, and $F_i$:

$$E(Y_i \mid X_{i1}, X_{i2}, F_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}^2 + \beta_4 X_{i2}^2 + \beta_5 X_{i1} X_{i2} + \delta_0 F_i . \quad (2.1)$$

- The ***female*** **population regression function**, or conditional mean function, implied by Model 2 is obtained by setting the female indicator variable $F_i = 1$ in (2.1):

$$E(Y_i \mid X_{i1}, X_{i2}, F_i = 1) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}^2 + \beta_4 X_{i2}^2 + \beta_5 X_{i1} X_{i2} + \delta_0 .$$

The ***female intercept*** **coefficient = $\beta_0 + \delta_0$**.

- The ***male*** **population regression function**, or conditional mean function, implied by Model 2 is obtained by setting the female indicator variable $F_i = 0$ in (2.1):

$$E(Y_i \mid X_{i1}, X_{i2}, F_i = 0) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}^2 + \beta_4 X_{i2}^2 + \beta_5 X_{i1} X_{i2} .$$

The ***male intercept*** **coefficient = $\beta_0$**.

- ***Interpretation* of the *female* indicator variable coefficient $\delta_0$:**

1. The slope coefficient $\delta_0$ of regressor $F_i$ in Model 2 equals the female intercept coefficient minus the male intercept coefficient:

   ***female* intercept coefficient – *male* intercept coefficient = $\beta_0 + \delta_0 - \beta_0 = \delta_0$**

2. A more substantive interpretation of $\delta_0$ can be obtained by subtracting the male population regression function $E(Y_i \mid X_{i1}, X_{i2}, F_i = 0)$ from the female population regression function $E(Y_i \mid X_{i1}, X_{i2}, F_i = 1)$:

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

   $$E(Y_i \mid X_{i1}, X_{i2}, F_i = 1) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}^2 + \beta_4 X_{i2}^2 + \beta_5 X_{i1} X_{i2} + \delta_0 .$$

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

   $$E(Y_i \mid X_{i1}, X_{i2}, F_i = 0) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}^2 + \beta_4 X_{i2}^2 + \beta_5 X_{i1} X_{i2} .$$

   The ***female-male difference* in mean Y** for given values of $X_1$ and $X_2$ is thus:

   $$E(Y_i \mid X_{i1}, X_{i2}, F_i = 1) - E(Y_i \mid X_{i1}, X_{i2}, F_i = 0) = \delta_0$$

   The **coefficient $\delta_0$** of the female indicator variable in Model 2 is therefore the difference between:

   (1) the **conditional mean of Y for *females*** with given values of $X_1$ and $X_2$

   and

   (2) the **conditional mean of Y for *males*** with the *same* values of $X_1$ and $X_2$.

   In other words, the **coefficient $\delta_0$** of the female indicator variable in Model 2 is the **difference in mean Y between *females* and *males* with *identical* values** of the explanatory variables $X_1$ and $X_2$.

## ❒ Indicator Variables as Multiplicative Regressors: Dummy Variable Interaction Terms

***Nature:*** When indicator (dummy) variables are introduced multiplicatively as additional regressors in linear regression models, they enter as dummy variable interaction terms -- that is, as the product of a dummy variable with some other variable, where the other variable may be either a continuous variable or another dummy variable.

There are therefore two types of dummy variable interaction terms.

1. **Interactions of a *dummy* variable with a *continuous* variable** -- that is, the product of a dummy variable and a continuous variable.
2. **Interactions of *one dummy* variable with *another dummy* variable --** that is, the product of one dummy variable and another dummy variable.

***Usage:*** ***Dummy variable interaction terms*** that equal the product of a ***continuous* variable** and an ***indicator* (*dummy*) variable** allow the slope coefficient of the continuous explanatory variable to differ between the two population subgroups identified by the indicator variable.

***Model 1:***

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}^2 + \beta_4 X_{i2}^2 + \beta_5 X_{i1} X_{i2} + u_i \qquad \textbf{(1)}$$

***Model 2:***

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}^2 + \beta_4 X_{i2}^2 + \beta_5 X_{i1} X_{i2} + \delta_0 F_i + u_i \qquad \textbf{(2)}$$

- Since both explanatory variables $X_1$ and $X_2$ in Models 1 and 2 are continuous variables, the five regressors $X_1$, $X_2$, $X_1^2$, $X_2^2$, and $X_1X_2$ are also continuous variables.

- To allow for different male and female slope coefficients on any of the five regressors $X_{i1}$, $X_{i2}$, $X_{i1}^2$, $X_{i2}^2$, and $X_{i1}X_{i2}$, add as additional regressors interaction terms between the female indicator variable $F_i$ and the continuous regressor.

  - To allow the slope coefficient of the regressor $X_{i1}$ to differ between females and males, add as an additional regressor to Model 1 or Model 2 the dummy variable interaction term $F_iX_{i1}$.

  - To allow the slope coefficient of the regressor $X_{i1}X_{i2}$ to differ between females and males, add as an additional regressor to Model 1 or Model 2 the dummy variable interaction term $F_iX_{i1}X_{i2}$.

  - To allow the slope coefficients of all five regressors $X_{i1}$, $X_{i2}$, $X_{i1}^2$, $X_{i2}^2$, and $X_{i1}X_{i2}$ to differ between females and males, add as additional regressors to Model 1 or Model 2 the five dummy variable interaction terms $F_iX_{i1}$, $F_iX_{i2}$, $F_iX_{i1}^2$, $F_iX_{i2}^2$, and $F_iX_{i1}X_{i2}$.

***<u>Model 3</u>:*** **a full-interaction regression equation**

Includes as regressors female dummy variable interaction terms with all five of the continuous regressors $X_{i1}$, $X_{i2}$, $X_{i1}^2$, $X_{i2}^2$, and $X_{i1}X_{i2}$.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}^2 + \beta_4 X_{i2}^2 + \beta_5 X_{i1}X_{i2} + \delta_0 F_i + \delta_1 F_i X_{i1} + \delta_2 F_i X_{i2} + \delta_3 F_i X_{i1}^2 + \delta_4 F_i X_{i2}^2 + \delta_5 F_i X_{i1} X_{i2} + u_i \qquad \textbf{(3)}$$

- The **population regression function**, or **conditional mean function**, **for Model 3** is obtained by taking the conditional expectation of regression equation (3) for any given values of the three explanatory variables $X_{i1}$, $X_{i2}$, and $F_i$:

$$E(Y_i | X_{i1}, X_{i2}, F_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}^2 + \beta_4 X_{i2}^2 + \beta_5 X_{i1}X_{i2} + \delta_0 F_i + \delta_1 F_i X_{i1} + \delta_2 F_i X_{i2} + \delta_3 F_i X_{i1}^2 + \delta_4 F_i X_{i2}^2 + \delta_5 F_i X_{i1} X_{i2} \qquad (3.1)$$

  - The ***female*** **regression function**, or female CMF, is obtained by setting the female indicator variable $\mathbf{F_i = 1}$ **in (3.1)**.
  - The ***male*** **regression function**, or male CMF, is obtained by setting the female indicator variable $\mathbf{F_i = 0}$ **in (3.1)**.

$$\begin{aligned} E(Y_i \mid X_{i1}, X_{i2}, F_i) = {} & \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}^2 + \beta_4 X_{i2}^2 + \beta_5 X_{i1} X_{i2} \\ & + \delta_0 F_i + \delta_1 F_i X_{i1} + \delta_2 F_i X_{i2} + \delta_3 F_i X_{i1}^2 + \delta_4 F_i X_{i2}^2 + \delta_5 F_i X_{i1} X_{i2} \end{aligned} \qquad (3.1)$$

- The ***female*** **population regression function**, or conditional mean function, implied by Model 3 is obtained by setting the female indicator variable $\mathbf{F_i = 1}$ **in (3.1)**:

$$\begin{aligned} E(Y_i \mid X_{i1}, X_{i2}, F_i = 1) = {} & \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}^2 + \beta_4 X_{i2}^2 + \beta_5 X_{i1} X_{i2} \\ & + \delta_0 + \delta_1 X_{i1} + \delta_2 X_{i2} + \delta_3 X_{i1}^2 + \delta_4 X_{i2}^2 + \delta_5 X_{i1} X_{i2} \\ = {} & (\beta_0 + \delta_0) + (\beta_1 + \delta_1) X_{i1} + (\beta_2 + \delta_2) X_{i2} \\ & + (\beta_3 + \delta_3) X_{i1}^2 + (\beta_4 + \delta_4) X_{i2}^2 + (\beta_5 + \delta_5) X_{i1} X_{i2} \\ = {} & \alpha_0 + \alpha_1 X_{i1} + \alpha_2 X_{i2} + \alpha_3 X_{i1}^2 + \alpha_4 X_{i2}^2 + \alpha_5 X_{i1} X_{i2} \end{aligned} \qquad (3.2)$$

where the female regression coefficients are $\alpha_j = \beta_j + \delta_j$ for all j = 0, 1, …, 5.

*female* intercept coefficient $= \alpha_0 = \beta_0 + \delta_0$
*female* slope coefficient of $X_{i1}$ $= \alpha_1 = \beta_1 + \delta_1$
*female* slope coefficient of $X_{i2}$ $= \alpha_2 = \beta_2 + \delta_2$
*female* slope coefficient of $X_{i1}^2$ $= \alpha_3 = \beta_3 + \delta_3$
*female* slope coefficient of $X_{i2}^2$ $= \alpha_4 = \beta_4 + \delta_4$
*female* slope coefficient of $X_{i1}X_{i2}$ $= \alpha_5 = \beta_5 + \delta_5$

- The ***male*** **population regression function**, or conditional mean function, implied by Model 3 is obtained by setting the female indicator variable $\mathbf{F_i = 0}$ **in (3.1)**:

$$E(Y_i \mid X_{i1}, X_{i2}, F_i = 0) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}^2 + \beta_4 X_{i2}^2 + \beta_5 X_{i1} X_{i2} \qquad (3.3)$$

*male* intercept coefficient $= \beta_0$
*male* slope coefficient of $X_{i1}$ $= \beta_1$
*male* slope coefficient of $X_{i2}$ $= \beta_2$
*male* slope coefficient of $X_{i1}^2$ $= \beta_3$
*male* slope coefficient of $X_{i2}^2$ $= \beta_4$
*male* slope coefficient of $X_{i1}X_{i2}$ $= \beta_5$

- The ***difference*** **between the** ***female*** **and** ***male*** **regression functions** -- that is, the ***female-male difference*** **in mean Y** for given (equal) values of the explanatory variables $X_1$ and $X_2$ -- is:

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

$$\begin{aligned} &= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}^2 + \beta_4 X_{i2}^2 + \beta_5 X_{i1} X_{i2} \\ &\quad + \delta_0 + \delta_1 X_{i1} + \delta_2 X_{i2} + \delta_3 X_{i1}^2 + \delta_4 X_{i2}^2 + \delta_5 X_{i1} X_{i2} \\ &\quad - \beta_0 - \beta_1 X_{i1} - \beta_2 X_{i2} - \beta_3 X_{i1}^2 - \beta_4 X_{i2}^2 - \beta_5 X_{i1} X_{i2} \\ &= \delta_0 + \delta_1 X_{i1} + \delta_2 X_{i2} + \delta_3 X_{i1}^2 + \delta_4 X_{i2}^2 + \delta_5 X_{i1} X_{i2} \end{aligned}$$

***Result:***

$$E(Y_i \mid X_{i1}, X_{i2}, F_i = 1) - E(Y_i \mid X_{i1}, X_{i2}, F_i = 0)$$
$$= \delta_0 + \delta_1 X_{i1} + \delta_2 X_{i2} + \delta_3 X_{i1}^2 + \delta_4 X_{i2}^2 + \delta_5 X_{i1} X_{i2}$$

***Interpretation:***

- The female-male difference in the conditional mean value of Y for given values $X_{i1}$ and $X_{i2}$ of the explanatory variables $X_1$ and $X_2$ is a quadratic function of $X_{i1}$ and $X_{i2}$. It is not a constant, but instead depends on the values of the explanatory variables $X_1$ and $X_2$.

- The female-male conditional mean Y difference addresses the following question: What is the female-male difference in mean Y for *identical* (*equal*) *values* of the explanatory variables $X_1$ and $X_2$.

- ***Interpretation*** **of the regression coefficients $\delta_j$ (j = 0, 1, …, 5) in Model 3**

  Each of the $\delta_j$ coefficients in Model 3 equals a ***female*** regression coefficient *minus* the corresponding ***male*** regression coefficient: $\delta_j = \alpha_j - \beta_j$ for all j.

  $\delta_0 = \alpha_0 - \beta_0$ = *female* intercept coefficient – *male* intercept coefficient
  $\delta_1 = \alpha_1 - \beta_1$ = *female* slope coefficient of $X_{i1}$ – *male* slope coefficient of $X_{i1}$
  $\delta_2 = \alpha_2 - \beta_2$ = *female* slope coefficient of $X_{i2}$ – *male* slope coefficient of $X_{i2}$
  $\delta_3 = \alpha_3 - \beta_3$ = *female* slope coefficient of $X_{i1}^2$ – *male* slope coefficient of $X_{i1}^2$
  $\delta_4 = \alpha_4 - \beta_4$ = *female* slope coefficient of $X_{i2}^2$ – *male* slope coefficient of $X_{i2}^2$
  $\delta_5 = \alpha_5 - \beta_5$ = *female* slope coefficient of $X_{i1}X_{i2}$ – *male* slope coefficient of $X_{i1}X_{i2}$

- The ***marginal*** **effects on Y of the two explanatory variables $X_1$ and $X_2$** in Model (3) are obtained by partially differentiating Y, or the conditional mean of Y given $X_1$ and $X_2$, with respect to each of the explanatory variables $X_1$ and $X_2$.

$$\begin{aligned} Y_i = {} & \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}^2 + \beta_4 X_{i2}^2 + \beta_5 X_{i1} X_{i2} \\ & + \delta_0 F_i + \delta_1 F_i X_{i1} + \delta_2 F_i X_{i2} + \delta_3 F_i X_{i1}^2 + \delta_4 F_i X_{i2}^2 + \delta_5 F_i X_{i1} X_{i2} + u_i \end{aligned} \quad \textbf{(3)}$$

$$\begin{aligned} E(Y_i \mid X_{i1}, X_{i2}, F_i) = {} & \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}^2 + \beta_4 X_{i2}^2 + \beta_5 X_{i1} X_{i2} \\ & + \delta_0 F_i + \delta_1 F_i X_{i1} + \delta_2 F_i X_{i2} + \delta_3 F_i X_{i1}^2 + \delta_4 F_i X_{i2}^2 + \delta_5 F_i X_{i1} X_{i2} \end{aligned} \quad (3.1)$$

**1.** The **marginal effect of $X_1$** in Model 3 is:

$$\frac{\partial Y_i}{\partial X_{i1}} = \frac{\partial E\left(Y_i \mid X_{i1}, X_{i2}, F_i\right)}{\partial X_{i1}}$$

$$= \beta_1 + 2\beta_3 X_{i1} + \beta_5 X_{i2} + \delta_1 F_i + 2\delta_3 F_i X_{i1} + \delta_5 F_i X_{i2}$$

**2.** The **marginal effect of $X_2$** in Model 3 is:

$$\frac{\partial Y_i}{\partial X_{i2}} = \frac{\partial E\left(Y_i \mid X_{i1}, X_{i2}\right)}{\partial X_{i2}}$$

$$= \beta_2 + 2\beta_4 X_{i2} + \beta_5 X_{i1} + \delta_2 F_i + 2\delta_4 F_i X_{i2} + \delta_5 F_i X_{i1}$$

- The **marginal effect of $X_1$** in Model 3 is ***different* for *males* and *females***.

$$\frac{\partial Y_i}{\partial X_{i1}} = \frac{\partial E(Y_i \mid X_{i1}, X_{i2}, F_i)}{\partial X_{i1}}$$

$$= \beta_1 + 2\beta_3 X_{i1} + \beta_5 X_{i2} + \delta_1 F_i + 2\delta_3 F_i X_{i1} + \delta_5 F_i X_{i2}$$

- ♦ The **marginal effect of $X_1$ for *males*** is obtained by setting $F_i = 0$ in the above equation:

$$\left(\frac{\partial Y_i}{\partial X_{i1}}\right)_{F_i=0} = \frac{\partial E(Y_i \mid X_{i1}, X_{i2}, F_i = 0)}{\partial X_{i1}}$$

$$= \beta_1 + 2\beta_3 X_{i1} + \beta_5 X_{i2}$$

- ♦ The **marginal effect of $X_1$ for *females*** is obtained by setting $F_i = 1$ in the above equation:

$$\left(\frac{\partial Y_i}{\partial X_{i1}}\right)_{F_i=1} = \frac{\partial E(Y_i \mid X_{i1}, X_{i2}, F_i = 1)}{\partial X_{i1}}$$

$$= \beta_1 + 2\beta_3 X_{i1} + \beta_5 X_{i2} + \delta_1 + 2\delta_3 X_{i1} + \delta_5 X_{i2}$$

$$= (\beta_1 + \delta_1) + 2(\beta_3 + \delta_3) X_{i1} + (\beta_5 + \delta_5) X_{i2}$$

$$= \alpha_1 + 2\alpha_3 X_{i1} + \alpha_5 X_{i2} \qquad \text{where } \alpha_j = \beta_j + \delta_j,\ j = 1, 3, 5$$

- The female-male difference in the **marginal effect of $X_1$** is obtained by subtracting the **marginal effect of $X_1$ for *males*** from the **marginal effect of $X_1$ for *females***:

$$\left(\frac{\partial Y_i}{\partial X_{i1}}\right)_{F_i=1} - \left(\frac{\partial Y_i}{\partial X_{i1}}\right)_{F_i=0} = \frac{\partial E\left(Y_i \mid X_{i1}, X_{i2}, F_i = 1\right)}{\partial X_{i1}} - \frac{\partial E\left(Y_i \mid X_{i1}, X_{i2}, F_i = 0\right)}{\partial X_{i1}}$$

$$= \beta_1 + 2\beta_3 X_{i1} + \beta_5 X_{i2} + \delta_1 + 2\delta_3 X_{i1} + \delta_5 X_{i2} - (\beta_1 + 2\beta_3 X_{i1} + \beta_5 X_{i2})$$

$$= \beta_1 + 2\beta_3 X_{i1} + \beta_5 X_{i2} + \delta_1 + 2\delta_3 X_{i1} + \delta_5 X_{i2} - \beta_1 - 2\beta_3 X_{i1} - \beta_5 X_{i2}$$

$$= \delta_1 + 2\delta_3 X_{i1} + \delta_5 X_{i2}$$

- The **marginal effect of $X_2$** in Model 3 is ***different* for *males* and *females***.

$$\frac{\partial Y_i}{\partial X_{i2}} = \frac{\partial E(Y_i \mid X_{i1}, X_{i2})}{\partial X_{i2}}$$

$$= \beta_2 + 2\beta_4 X_{i2} + \beta_5 X_{i1} + \delta_2 F_i + 2\delta_4 F_i X_{i2} + \delta_5 F_i X_{i1}$$

- The **marginal effect of $X_2$ for *males*** is obtained by setting $F_i = 0$ in the above equation:

$$\left(\frac{\partial Y_i}{\partial X_{i2}}\right)_{F_i=0} = \frac{\partial E(Y_i \mid X_{i1}, X_{i2}, F_i = 0)}{\partial X_{i2}}$$

$$= \beta_2 + 2\beta_4 X_{i2} + \beta_5 X_{i1}$$

- The **marginal effect of $X_2$ for *females*** is obtained by setting $F_i = 1$ in the above equation:

$$\left(\frac{\partial Y_i}{\partial X_{i2}}\right)_{F_i=1} = \frac{\partial E(Y_i \mid X_{i1}, X_{i2}, F_i = 1)}{\partial X_{i2}}$$

$$= \beta_2 + 2\beta_4 X_{i2} + \beta_5 X_{i1} + \delta_2 + 2\delta_4 X_{i2} + \delta_5 X_{i1}$$

$$= (\beta_2 + \delta_2) + 2(\beta_4 + \delta_4) X_{i2} + (\beta_5 + \delta_5) X_{i1}$$

$$= \alpha_2 + 2\alpha_4 X_{i2} + \alpha_5 X_{i1} \qquad \text{where } \alpha_j = \beta_j + \delta_j,\ j = 2, 4, 5$$

- The female-male difference in the **marginal effect of $X_2$** is obtained by subtracting the **marginal effect of $X_2$ for *males*** from the **marginal effect of $X_2$ for *females***:

$$\left(\frac{\partial Y_i}{\partial X_{i2}}\right)_{F_i=1} - \left(\frac{\partial Y_i}{\partial X_{i2}}\right)_{F_i=0} = \frac{\partial E\left(Y_i \,|\, X_{i1}, X_{i2}, F_i = 1\right)}{\partial X_{i2}} - \frac{\partial E\left(Y_i \,|\, X_{i1}, X_{i2}, F_i = 0\right)}{\partial X_{i2}}$$

$$= \beta_2 + 2\beta_4 X_{i2} + \beta_5 X_{i1} + \delta_2 + 2\delta_4 X_{i2} + \delta_5 X_{i1} - (\beta_2 + 2\beta_4 X_{i2} + \beta_5 X_{i1})$$

$$= \beta_2 + 2\beta_4 X_{i2} + \beta_5 X_{i1} + \delta_2 + 2\delta_4 X_{i2} + \delta_5 X_{i1} - \beta_2 - 2\beta_4 X_{i2} - \beta_5 X_{i1}$$

$$= \delta_2 + 2\delta_4 X_{i2} + \delta_5 X_{i1}$$

## ❐ Tests for Female-Male Coefficient Differences in Model 3

Re-write the population regression equation and population regression function for Model 3:

$$\begin{aligned} Y_i = {} & \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}^2 + \beta_4 X_{i2}^2 + \beta_5 X_{i1} X_{i2} \\ & + \delta_0 F_i + \delta_1 F_i X_{i1} + \delta_2 F_i X_{i2} + \delta_3 F_i X_{i1}^2 + \delta_4 F_i X_{i2}^2 + \delta_5 F_i X_{i1} X_{i2} + u_i \end{aligned} \qquad \textbf{(3)}$$

$$\begin{aligned} E(Y_i \mid X_{i1}, X_{i2}, F_i) = {} & \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}^2 + \beta_4 X_{i2}^2 + \beta_5 X_{i1} X_{i2} \\ & + \delta_0 F_i + \delta_1 F_i X_{i1} + \delta_2 F_i X_{i2} + \delta_3 F_i X_{i1}^2 + \delta_4 F_i X_{i2}^2 + \delta_5 F_i X_{i1} X_{i2} \end{aligned} \qquad (3.1)$$

Any hypothesis about coefficient differences between males and females can be formulated as restrictions on the $\delta_j$ regression coefficients in Model 3, each of which is equal to a female regression coefficient minus the corresponding male regression coefficient.

**$\delta_j$ = *female* coefficient of regressor j – *male* coefficient of regressor j**

This section gives several examples of hypotheses that can be formulated as restrictions on the $\delta_j$ coefficients in Model 3.

- ***Test 1:*** Test the proposition that **males and females have *identical* mean values of Y for any given values of $X_1$ and $X_2$**.

- Recall that the female-male difference in the conditional mean value of Y for any specified values of $X_1$ and $X_2$ is given in Model 3 by

$$E(Y_i \mid X_{i1}, X_{i2}, F_i = 1) - E(Y_i \mid X_{i1}, X_{i2}, F_i = 0)$$
$$= \delta_0 + \delta_1 X_{i1} + \delta_2 X_{i2} + \delta_3 X_{i1}^2 + \delta_4 X_{i2}^2 + \delta_5 X_{i1} X_{i2}$$

- The proposition to be tested is that

$$E(Y_i \mid X_{i1}, X_{i2}, F_i = 1) = E(Y_i \mid X_{i1}, X_{i2}, F_i = 0) \qquad \text{for all i}$$

which implies that

$$E(Y_i \mid X_{i1}, X_{i2}, F_i = 1) - E(Y_i \mid X_{i1}, X_{i2}, F_i = 0) = 0 \qquad \text{for all i}$$

and hence that

$$\delta_0 + \delta_1 X_{i1} + \delta_2 X_{i2} + \delta_3 X_{i1}^2 + \delta_4 X_{i2}^2 + \delta_5 X_{i1} X_{i2} = 0 \qquad \text{for all i}$$

- A ***sufficient* condition** for these statements to be true is that ***all six* of the $\delta_j$ coefficients** in Model 3 jointly **equal *zero***.

- The ***null*** **and** ***alternative*** **hypotheses** are as follows:

  $H_0$: $\delta_j = 0$ for all j = 0, 1, …, 5

  $\delta_0 = 0$ *and* $\delta_1 = 0$ *and* $\delta_2 = 0$ *and* $\delta_3 = 0$ *and* $\delta_4 = 0$ *and* $\delta_5 = 0$

  $H_1$: $\delta_j \neq 0$ j = 0, 1, …, 5

  $\delta_0 \neq 0$ *and/or* $\delta_1 \neq 0$ *and/or* $\delta_2 \neq 0$ *and/or* $\delta_3 \neq 0$ *and/or* $\delta_4 \neq 0$ *and/or* $\delta_5 \neq 0$

- The ***restricted*** **model implied by the null hypothesis $H_0$** is obtained by imposing on Model 3 (the unrestricted model) the coefficient restrictions specified by $H_0$.

  ***Model 3, the unrestricted model,*** is:

$$
\begin{aligned}
Y_i = {} & \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}^2 + \beta_4 X_{i2}^2 + \beta_5 X_{i1} X_{i2} \\
& + \delta_0 F_i + \delta_1 F_i X_{i1} + \delta_2 F_i X_{i2} + \delta_3 F_i X_{i1}^2 + \delta_4 F_i X_{i2}^2 + \delta_5 F_i X_{i1} X_{i2} + u_i
\end{aligned} \qquad \textbf{(3)}
$$

  The ***restricted*** **model** is obtained by setting $\delta_j = 0$ for all j = 0, 1, …, 5 in Model 3:

$$
Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}^2 + \beta_4 X_{i2}^2 + \beta_5 X_{i1} X_{i2} + u_i \qquad \textbf{(1)}
$$

- The ***test statistic*** appropriate for this hypothesis test is a **Wald F-statistic**.

- ***Test 2:*** Test the proposition that the **female-male difference in mean Y is a *constant***, i.e., that it does not depend on the explanatory variables $X_1$ and $X_2$.

- Recall that the female-male difference in the conditional mean value of Y for any specified values of $X_1$ and $X_2$ is given in Model 3 by

$$E(Y_i \mid X_{i1}, X_{i2}, F_i = 1) - E(Y_i \mid X_{i1}, X_{i2}, F_i = 0) = \delta_0 + \delta_1 X_{i1} + \delta_2 X_{i2} + \delta_3 X_{i1}^2 + \delta_4 X_{i2}^2 + \delta_5 X_{i1} X_{i2}$$

- The hypothesis to be tested is that

$$E(Y_i \mid X_{i1}, X_{i2}, F_i = 1) - E(Y_i \mid X_{i1}, X_{i2}, F_i = 0) = \text{a constant} \quad \text{for all i}$$

which implies that

$$E(Y_i \mid X_{i1}, X_{i2}, F_i = 1) - E(Y_i \mid X_{i1}, X_{i2}, F_i = 0) = \delta_0 \quad \text{for all i}$$

and hence that

$$\delta_0 + \delta_1 X_{i1} + \delta_2 X_{i2} + \delta_3 X_{i1}^2 + \delta_4 X_{i2}^2 + \delta_5 X_{i1} X_{i2} = \delta_0 \quad \text{for all i}$$

- A ***sufficient* condition** for these statements to be true is that **the *five* $\delta_j$ coefficients** on the female dummy variable interaction terms in Model 3 **all equal *zero***.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}^2 + \beta_4 X_{i2}^2 + \beta_5 X_{i1} X_{i2} + \delta_0 F_i + \delta_1 F_i X_{i1} + \delta_2 F_i X_{i2} + \delta_3 F_i X_{i1}^2 + \delta_4 F_i X_{i2}^2 + \delta_5 F_i X_{i1} X_{i2} + u_i \quad \textbf{(3)}$$

- The ***null*** **and** ***alternative*** **hypotheses** are as follows:

$H_0$: $\delta_j = 0$ for all j = 1, …, 5

$\delta_1 = 0$ *and* $\delta_2 = 0$ *and* $\delta_3 = 0$ *and* $\delta_4 = 0$ *and* $\delta_5 = 0$

$H_1$: $\delta_j \neq 0$ j = 1, …, 5

$\delta_1 \neq 0$ *and/or* $\delta_2 \neq 0$ *and/or* $\delta_3 \neq 0$ *and/or* $\delta_4 \neq 0$ *and/or* $\delta_5 \neq 0$

- The ***restricted*** **model implied by the null hypothesis $H_0$** is obtained by imposing on Model 3 (the unrestricted model) the coefficient restrictions specified by $H_0$.

***Model 3, the unrestricted model,*** is:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}^2 + \beta_4 X_{i2}^2 + \beta_5 X_{i1} X_{i2} + \delta_0 F_i + \delta_1 F_i X_{i1} + \delta_2 F_i X_{i2} + \delta_3 F_i X_{i1}^2 + \delta_4 F_i X_{i2}^2 + \delta_5 F_i X_{i1} X_{i2} + u_i \quad \textbf{(3)}$$

The ***restricted*** **model** is obtained by setting $\delta_j = 0$ for all j = 1, …, 5 in Model 3:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}^2 + \beta_4 X_{i2}^2 + \beta_5 X_{i1} X_{i2} + \delta_0 F_i + u_i \quad \textbf{(2)}$$

- The ***test statistic*** appropriate for this hypothesis test is a **Wald F-statistic**.

- ***Test 3:*** Test the proposition that the **female-male *difference* in mean Y does not depend on the explanatory variable $X_1$**.

  This proposition is empirically equivalent to the following three statements:

  **(1)** The relationship of Y to $X_1$ is identical for males and females.

  **(2)** The marginal effect of $X_1$ on Y is identical for males and females.

  **(3)** The female-male difference in mean Y is a function only of the explanatory variable $X_2$.

- Recall that the female-male difference in the conditional mean value of Y for any specified values of $X_1$ and $X_2$ is given in Model 3 by

$$E(Y_i \mid X_{i1}, X_{i2}, F_i = 1) - E(Y_i \mid X_{i1}, X_{i2}, F_i = 0)$$
$$= \delta_0 + \delta_1 X_{i1} + \delta_2 X_{i2} + \delta_3 X_{i1}^2 + \delta_4 X_{i2}^2 + \delta_5 X_{i1} X_{i2}$$

  The female-male difference in mean Y does not depend on $X_1$ if and only if $\boldsymbol{\delta_1 = 0}$ ***and*** $\boldsymbol{\delta_3 = 0}$ ***and*** $\boldsymbol{\delta_5 = 0}$. Under these three exclusion restrictions,

$$E(Y_i \mid X_{i1}, X_{i2}, F_i = 1) - E(Y_i \mid X_{i1}, X_{i2}, F_i = 0) = \delta_0 + \delta_2 X_{i2} + \delta_4 X_{i2}^2$$

- Recall that the **marginal effects of $X_1$ for *males* and *females*** in Model 3 are given respectively by:

Males: $$\frac{\partial E(Y_i \mid X_{i1}, X_{i2}, F_i = 0)}{\partial X_{i1}} = \beta_1 + 2\beta_3 X_{i1} + \beta_5 X_{i2}$$

Females: $$\frac{\partial E(Y_i \mid X_{i1}, X_{i2}, F_i = 1)}{\partial X_{i1}} = \beta_1 + 2\beta_3 X_{i1} + \beta_5 X_{i2} + \delta_1 + 2\delta_3 X_{i1} + \delta_5 X_{i2}$$

These two functions are identical (for any given values of $X_1$ and $X_2$) if and only if **$\delta_1 = 0$ *and* $\delta_3 = 0$ *and* $\delta_5 = 0$**.

- The ***null* and *alternative* hypotheses** are therefore as follows:

$H_0$: $\delta_j = 0$ for j = 1, 3, 5

$\delta_1 = 0$ *and* $\delta_3 = 0$ *and* $\delta_5 = 0$

$H_1$: $\delta_j \neq 0$ j = 1, 3, 5

$\delta_1 \neq 0$ *and/or* $\delta_3 \neq 0$ *and/or* $\delta_5 \neq 0$

- The ***restricted* model implied by the null hypothesis $H_0$** is obtained by imposing on Model 3 (the unrestricted model) the coefficient restrictions specified by $H_0$.

  ***Model 3, the unrestricted model,*** is:

$$\begin{aligned} Y_i = {} & \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}^2 + \beta_4 X_{i2}^2 + \beta_5 X_{i1} X_{i2} \\ & + \delta_0 F_i + \delta_1 F_i X_{i1} + \delta_2 F_i X_{i2} + \delta_3 F_i X_{i1}^2 + \delta_4 F_i X_{i2}^2 + \delta_5 F_i X_{i1} X_{i2} + u_i \end{aligned} \qquad \textbf{(3)}$$

  The ***restricted* model** is obtained by setting $\delta_1 = 0$, $\delta_3 = 0$, *and* $\delta_5 = 0$ in Model 3:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}^2 + \beta_4 X_{i2}^2 + \beta_5 X_{i1} X_{i2} + \delta_0 F_i + \delta_2 F_i X_{i2} + \delta_4 F_i X_{i2}^2 + u_i$$

- The ***test statistic*** appropriate for this hypothesis test is a **Wald F-statistic**.

- ***Test 4:*** Test the proposition that the **female-male *difference* in mean Y does not depend on the explanatory variable $X_2$**.

  This proposition is empirically equivalent to the following three statements:

  **(1)** The relationship of Y to $X_2$ is identical for males and females.

  **(2)** The marginal effect of $X_2$ on Y is identical for males and females.

  **(3)** The female-male difference in mean Y is a function only of the explanatory variable $X_1$.

- Recall that the female-male difference in the conditional mean value of Y for any specified values of $X_1$ and $X_2$ is given in Model 3 by

$$
\begin{aligned}
E(Y_i \mid X_{i1}, X_{i2}, F_i = 1) - E(Y_i \mid X_{i1}, X_{i2}, F_i = 0) \\
= \delta_0 + \delta_1 X_{i1} + \delta_2 X_{i2} + \delta_3 X_{i1}^2 + \delta_4 X_{i2}^2 + \delta_5 X_{i1} X_{i2}
\end{aligned}
$$

The female-male difference in mean Y does not depend on $X_2$ if and only if $\boldsymbol{\delta_2 = 0}$ ***and*** $\boldsymbol{\delta_4 = 0}$ ***and*** $\boldsymbol{\delta_5 = 0}$. Under these three exclusion restrictions,

$$E(Y_i \mid X_{i1}, X_{i2}, F_i = 1) - E(Y_i \mid X_{i1}, X_{i2}, F_i = 0) = \delta_0 + \delta_1 X_{i1} + \delta_3 X_{i1}^2$$

- Recall that the **marginal effects of $X_2$ for *males* and *females*** in Model 3 are given respectively by:

Males: $$\frac{\partial E(Y_i \mid X_{i1}, X_{i2}, F_i = 0)}{\partial X_{i2}} = \beta_2 + 2\beta_4 X_{i2} + \beta_5 X_{i1}$$

Females: $$\frac{\partial E(Y_i \mid X_{i1}, X_{i2}, F_i = 1)}{\partial X_{i2}} = \beta_2 + 2\beta_4 X_{i2} + \beta_5 X_{i1} + \delta_2 + 2\delta_4 X_{i2} + \delta_5 X_{i1}$$

These two functions are identical (for any given values of $X_1$ and $X_2$) if and only if **$\delta_2 = 0$ *and* $\delta_4 = 0$ *and* $\delta_5 = 0$**.

- The ***null* and *alternative* hypotheses** are therefore as follows:

$H_0$: $\delta_j = 0$ for j = 2, 4, 5

$\delta_2 = 0$ *and* $\delta_4 = 0$ *and* $\delta_5 = 0$

$H_1$: $\delta_j \neq 0$ j = 2, 4, 5

$\delta_2 \neq 0$ *and/or* $\delta_4 \neq 0$ *and/or* $\delta_5 \neq 0$

- The ***restricted*** **model implied by the null hypothesis $H_0$** is obtained by imposing on Model 3 (the unrestricted model) the coefficient restrictions specified by $H_0$.

  ***Model 3, the unrestricted model,*** is:

$$\begin{aligned} Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}^2 + \beta_4 X_{i2}^2 + \beta_5 X_{i1} X_{i2} \\ + \delta_0 F_i + \delta_1 F_i X_{i1} + \delta_2 F_i X_{i2} + \delta_3 F_i X_{i1}^2 + \delta_4 F_i X_{i2}^2 + \delta_5 F_i X_{i1} X_{i2} + u_i \end{aligned} \quad \textbf{(3)}$$

  The ***restricted*** **model** is obtained by setting $\delta_2 = 0$, $\delta_4 = 0$, *and* $\delta_5 = 0$ in Model 3:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}^2 + \beta_4 X_{i2}^2 + \beta_5 X_{i1} X_{i2} + \delta_0 F_i + \delta_1 F_i X_{i1} + \delta_3 F_i X_{i1}^2 + u_i$$

- The ***test statistic*** appropriate for this hypothesis test is a **Wald F-statistic**.

♦ ***Test 5:*** Test the proposition that the **female-male *difference* in mean Y is a *linear* function of the explanatory variables $X_1$ and $X_2$**.

- Recall that the female-male difference in the conditional mean value of Y for any specified values of $X_1$ and $X_2$ is given in Model 3 by

$$E(Y_i \mid X_{i1}, X_{i2}, F_i = 1) - E(Y_i \mid X_{i1}, X_{i2}, F_i = 0)$$
$$= \delta_0 + \delta_1 X_{i1} + \delta_2 X_{i2} + \delta_3 X_{i1}^2 + \delta_4 X_{i2}^2 + \delta_5 X_{i1} X_{i2}$$

The female-male difference in mean Y is linear in $X_1$ and $X_2$ if and only if $\delta_3 = 0$ ***and*** $\delta_4 = 0$ ***and*** $\delta_5 = 0$. Under these three exclusion restrictions,

$$E(Y_i \mid X_{i1}, X_{i2}, F_i = 1) - E(Y_i \mid X_{i1}, X_{i2}, F_i = 0) = \delta_0 + \delta_1 X_{i1} + \delta_2 X_{i2}$$

- Note the implications of the three coefficient restrictions **$\delta_3 = 0$, $\delta_4 = 0$ and $\delta_5 = 0$** for the **marginal effects of $X_1$ and $X_2$ for *females*** in Model 3, which are given respectively by:

Females: $$\frac{\partial E\left(Y_i \mid X_{i1}, X_{i2}, F_i = 1\right)}{\partial X_{i1}} = \beta_1 + 2\beta_3 X_{i1} + \beta_5 X_{i2} + \delta_1 + 2\delta_3 X_{i1} + \delta_5 X_{i2}$$

Females: $$\frac{\partial E\left(Y_i \mid X_{i1}, X_{i2}, F_i = 1\right)}{\partial X_{i2}} = \beta_2 + 2\beta_4 X_{i2} + \beta_5 X_{i1} + \delta_2 + 2\delta_4 X_{i2} + \delta_5 X_{i1}$$

**Under the coefficient restrictions $\delta_3 = 0$ and $\delta_4 = 0$ and $\delta_5 = 0$**, the **marginal effects of $X_1$ and $X_2$ for *females*** are:

Females: $$\frac{\partial E(Y_i \mid X_{i1}, X_{i2}, F_i = 1)}{\partial X_{i1}} = \beta_1 + 2\beta_3 X_{i1} + \beta_5 X_{i2} + \delta_1$$

Females: $$\frac{\partial E(Y_i \mid X_{i1}, X_{i2}, F_i = 1)}{\partial X_{i2}} = \beta_2 + 2\beta_4 X_{i2} + \beta_5 X_{i1} + \delta_2$$

In other words, under the coefficient restrictions $\delta_3 = 0$ and $\delta_4 = 0$ and $\delta_5 = 0$, the marginal effects of $X_1$ and $X_2$ for *females* differ from the marginal effects of $X_1$ and $X_2$ for *males* only by a constant.

$$\delta_1 = \frac{\partial E(Y_i \mid X_{i1}, X_{i2}, F_i = 1)}{\partial X_{i1}} - \frac{\partial E(Y_i \mid X_{i1}, X_{i2}, F_i = 0)}{\partial X_{i1}}$$

= marginal effect of $X_1$ for *females* – marginal effect of $X_1$ for *males*

$$\delta_2 = \frac{\partial E(Y_i \mid X_{i1}, X_{i2}, F_i = 1)}{\partial X_{i2}} - \frac{\partial E(Y_i \mid X_{i1}, X_{i2}, F_i = 0)}{\partial X_{i2}}$$

= marginal effect of $X_2$ for *females* – marginal effect of $X_2$ for *males*

- The ***null*** **and** ***alternative*** **hypotheses** are therefore as follows:

  $H_0$: $\delta_j = 0$ for j = 3, 4, 5

  $\delta_3 = 0$ *and* $\delta_4 = 0$ *and* $\delta_5 = 0$

  $H_1$: $\delta_j \neq 0$ j = 3, 4, 5

  $\delta_3 \neq 0$ *and/or* $\delta_4 \neq 0$ *and/or* $\delta_5 \neq 0$

- The ***restricted*** **model implied by the null hypothesis $H_0$** is obtained by imposing on Model 3 (the unrestricted model) the coefficient restrictions specified by $H_0$.

  ***Model 3, the unrestricted model,*** is:

$$\begin{aligned} Y_i = {} & \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}^2 + \beta_4 X_{i2}^2 + \beta_5 X_{i1} X_{i2} \\ & + \delta_0 F_i + \delta_1 F_i X_{i1} + \delta_2 F_i X_{i2} + \delta_3 F_i X_{i1}^2 + \delta_4 F_i X_{i2}^2 + \delta_5 F_i X_{i1} X_{i2} + u_i \end{aligned} \qquad \textbf{(3)}$$

  The ***restricted*** **model** is obtained by setting $\delta_3 = 0$, $\delta_4 = 0$, *and* $\delta_5 = 0$ in Model 3:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}^2 + \beta_4 X_{i2}^2 + \beta_5 X_{i1} X_{i2} + \delta_0 F_i + \delta_1 F_i X_{i1} + \delta_2 F_i X_{i2} + u_i$$

- The ***test statistic*** appropriate for this hypothesis test is a **Wald F-statistic**.

## ❒ Tests on the Marginal Effects of $X_1$ and $X_2$ for Males in Model 3

### Model 3: Tests to Perform on the Marginal Effect of $X_1$ for Males

The **marginal effect of $X_1$ for *males*** in Model 3 is:

$$\left(\frac{\partial Y_i}{\partial X_{i1}}\right)_{F_i=0} = \frac{\partial E\left(Y_i \mid X_{i1}, X_{i2}, F_i = 0\right)}{\partial X_{i1}} = \beta_1 + 2\beta_3 X_{i1} + \beta_5 X_{i2}$$

- **Test 1m:** Test the proposition that the **marginal effect of $X_1$ for *males* is *zero*** for any values of the two continuous explanatory variables $X_1$ and $X_2$.

  $H_0$: $\beta_j = 0$ for j = 1, 3, 5

  $\beta_1 = 0$ *and* $\beta_3 = 0$ *and* $\beta_5 = 0$

  $H_1$: $\beta_j \neq 0$ j = 1, 3, 5

  $\beta_1 \neq 0$ *and/or* $\beta_3 \neq 0$ *and/or* $\beta_5 \neq 0$

  Perform an **F-test** of these **three** coefficient exclusion restrictions using the ***Stata* test** command.

The **marginal effect of $X_1$ for *males*** in Model 3 is:

$$\left(\frac{\partial Y_i}{\partial X_{i1}}\right)_{F_i=0} = \frac{\partial E\left(Y_i \mid X_{i1}, X_{i2}, F_i = 0\right)}{\partial X_{i1}} = \beta_1 + 2\beta_3 X_{i1} + \beta_5 X_{i2}$$

- **<u>Test 2m</u>:** Test the proposition that the **marginal effect of $X_1$ for *males* is a *constant*,** i.e., that is does not depend upon the values of $X_1$ or $X_2$.

  $H_0$: $\beta_j = 0$ for j = 3, 5

  $\beta_3 = 0$ *and* $\beta_5 = 0$

  $H_1$: $\beta_j \neq 0$ j = 3, 5

  $\beta_3 \neq 0$ *and/or* $\beta_5 \neq 0$

  Perform an **F-test** of these **two** coefficient exclusion restrictions using the ***Stata* test** command .

The **marginal effect of $X_1$ for *males*** in Model 3 is:

$$\left(\frac{\partial Y_i}{\partial X_{i1}}\right)_{F_i=0} = \frac{\partial E\left(Y_i \mid X_{i1}, X_{i2}, F_i = 0\right)}{\partial X_{i1}} = \beta_1 + 2\beta_3 X_{i1} + \beta_5 X_{i2}$$

- **<u>Test 3m</u>:** Test the proposition that the **marginal effect of $X_1$ for *males* does not depend upon**, or is unrelated to, **the value of $X_2$**.

  $H_0$: $\beta_5 = 0$

  $H_1$: $\beta_5 \neq 0$

  Perform ***either*** an **F-test *or*** a **two-tail t-test** of this **one** coefficient exclusion restriction.

- **<u>Test 4m</u>:** Test the proposition that the **marginal effect of $X_1$ for *males* does not depend upon**, or is unrelated to, **the value of $X_1$**.

  $H_0$: $\beta_3 = 0$

  $H_1$: $\beta_3 \neq 0$

  Perform ***either*** an **F-test *or*** a **two-tail t-test** of this **one** coefficient exclusion restriction.

**Model 3: Tests to Perform on the Marginal Effect of $X_2$ for Males**

Formulate the analogs of Tests 1m to 4m for the **marginal effect of $X_2$ for *males***, which is

$$\left(\frac{\partial Y_i}{\partial X_{i2}}\right)_{F_i=0} = \frac{\partial E(Y_i \mid X_{i1}, X_{i2}, F_i = 0)}{\partial X_{i2}} = \beta_2 + 2\beta_4 X_{i2} + \beta_5 X_{i1}$$

## ❒ Tests on the Marginal Effects of $X_1$ and $X_2$ for Females in Model 3

### <u>Model 3</u>: Tests to Perform on the Marginal Effect of $X_1$ for Females

The **marginal effect of $X_1$ for *females*** in Model 3 is:

$$\left(\frac{\partial Y_i}{\partial X_{i1}}\right)_{F_i=1} = \frac{\partial E\left(Y_i \mid X_{i1}, X_{i2}, F_i = 1\right)}{\partial X_{i1}} = \beta_1 + 2\beta_3 X_{i1} + \beta_5 X_{i2} + \delta_1 + 2\delta_3 X_{i1} + \delta_5 X_{i2}$$

$$= (\beta_1 + \delta_1) + 2(\beta_3 + \delta_3) X_{i1} + (\beta_5 + \delta_5) X_{i2}$$

- **<u>Test 1f</u>:** Test the proposition that the **marginal effect of $X_1$ for *females* is *zero* for *any* values** of the two continuous explanatory variables $X_1$ and $X_2$.

  $H_0$: $\beta_j + \delta_j = 0$ for j = 1, 3, 5

  $\beta_1 + \delta_1 = 0$ *and* $\beta_3 + \delta_3 = 0$ *and* $\beta_5 + \delta_5 = 0$

  $H_1$: $\beta_j + \delta_j \neq 0$ j = 1, 3, 5

  $\beta_1 + \delta_1 \neq 0$ *and/or* $\beta_3 + \delta_3 \neq 0$ *and/or* $\beta_5 + \delta_5 \neq 0$

  Perform an **F-test** of these **three** coefficient exclusion restrictions; use a sequence of three ***Stata* test** commands with the **accumulate** option.

```
test x1 + fx1 = 0, notest
test x1sq + fx1sq = 0, notest accumulate
test x1x2 + fx1x2 = 0, accumulate
```

The **marginal effect of $X_1$ for *females*** in Model 3 is:

$$\left(\frac{\partial Y_i}{\partial X_{i1}}\right)_{F_i=1} = \frac{\partial E(Y_i \mid X_{i1}, X_{i2}, F_i = 1)}{\partial X_{i1}} = \beta_1 + 2\beta_3 X_{i1} + \beta_5 X_{i2} + \delta_1 + 2\delta_3 X_{i1} + \delta_5 X_{i2}$$

$$= (\beta_1 + \delta_1) + 2(\beta_3 + \delta_3)X_{i1} + (\beta_5 + \delta_5)X_{i2}$$

- **Test 2f:** Test the proposition that the **marginal effect of $X_1$ for *females* is a *constant*,** i.e., that is does not depend upon the values of $X_1$ or $X_2$.

  $H_0$: $\beta_j + \delta_j = 0$ for j = 3, 5

  $\beta_3 + \delta_3 = 0$ *and* $\beta_5 + \delta_5 = 0$

  $H_1$: $\beta_j + \delta_j \neq 0$ j = 3, 5

  $\beta_3 + \delta_3 \neq 0$ *and/or* $\beta_5 + \delta_5 \neq 0$

  Perform an **F-test** of these **two** coefficient exclusion restrictions; use a sequence of two ***Stata* test** commands with the **accumulate** option.

```
test x1sq + fx1sq = 0, notest
test x1x2 + fx1x2 = 0, accumulate
```

The **marginal effect of $X_1$ for *females*** in Model 3 is:

$$\left(\frac{\partial Y_i}{\partial X_{i1}}\right)_{F_i=1} = \frac{\partial E(Y_i \mid X_{i1}, X_{i2}, F_i = 1)}{\partial X_{i1}}$$

$$= \beta_1 + 2\beta_3 X_{i1} + \beta_5 X_{i2} + \delta_1 + 2\delta_3 X_{i1} + \delta_5 X_{i2}$$

$$= (\beta_1 + \delta_1) + 2(\beta_3 + \delta_3)X_{i1} + (\beta_5 + \delta_5)X_{i2}$$

- **Test 3f:** Test the proposition that the **marginal effect of $X_1$ for *females* does not depend upon**, or is unrelated to, **the value of $X_2$**.

  $H_0$: $\beta_5 + \delta_5 = 0$

  $H_1$: $\beta_5 + \delta_5 \neq 0$

  Perform ***either*** an **F-test *or*** a **two-tail t-test** of this **one** coefficient exclusion restriction.

- **Test 4f:** Test the proposition that the **marginal effect of $X_1$ for *females* does not depend upon**, or is unrelated to, **the value of $X_1$**.

  $H_0$: $\beta_3 + \delta_3 = 0$

  $H_1$: $\beta_3 + \delta_3 \neq 0$

  Perform ***either*** an **F-test *or*** a **two-tail t-test** of this **one** coefficient exclusion restriction.

**<u>Model 3</u>: Tests to Perform on the Marginal Effect of $X_2$ for Females**

Formulate the analogs of Tests 1f to 4f for the **marginal effect of $X_2$ for *females***, which is

$$\left(\frac{\partial Y_i}{\partial X_{i2}}\right)_{F_i=1} = \frac{\partial E\left(Y_i \mid X_{i1}, X_{i2}, F_i = 1\right)}{\partial X_{i2}}$$

$$= \beta_2 + 2\beta_4 X_{i2} + \beta_5 X_{i1} + \delta_2 + 2\delta_4 X_{i2} + \delta_5 X_{i1}$$

$$= (\beta_2 + \delta_2) + 2(\beta_4 + \delta_4) X_{i2} + (\beta_5 + \delta_5) X_{i1}$$

## ❒ An Alternative Formulation of Model 3 Using the Male Dummy Variable $M_i$

***<u>Model 3</u>:* a full-interaction regression equation** in the ***female* dummy variable $F_i$**

Recall that the **population regression *equation*** and **population regression *function*** for **Model 3** are:

$$\begin{aligned} Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}^2 + \beta_4 X_{i2}^2 + \beta_5 X_{i1} X_{i2} \\ + \delta_0 F_i + \delta_1 F_i X_{i1} + \delta_2 F_i X_{i2} + \delta_3 F_i X_{i1}^2 + \delta_4 F_i X_{i2}^2 + \delta_5 F_i X_{i1} X_{i2} + u_i \end{aligned} \quad \textbf{(3)}$$

$$\begin{aligned} E(Y_i | X_{i1}, X_{i2}, F_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}^2 + \beta_4 X_{i2}^2 + \beta_5 X_{i1} X_{i2} \\ + \delta_0 F_i + \delta_1 F_i X_{i1} + \delta_2 F_i X_{i2} + \delta_3 F_i X_{i1}^2 + \delta_4 F_i X_{i2}^2 + \delta_5 F_i X_{i1} X_{i2} \end{aligned} \quad (3.1)$$

***<u>Model 3*</u>:* an alternative full-interaction regression equation** in the ***male* dummy variable $M_i$**

- The **population regression equation for Model 3*** is

$$\begin{aligned} Y_i = \alpha_0 + \alpha_1 X_{i1} + \alpha_2 X_{i2} + \alpha_3 X_{i1}^2 + \alpha_4 X_{i2}^2 + \alpha_5 X_{i1} X_{i2} \\ + \gamma_0 M_i + \gamma_1 M_i X_{i1} + \gamma_2 M_i X_{i2} + \gamma_3 M_i X_{i1}^2 + \gamma_4 M_i X_{i2}^2 + \gamma_5 M_i X_{i1} X_{i2} + u_i \end{aligned} \quad \textbf{(3*)}$$

- The **population regression function**, or **conditional mean function**, **for Model 3*** is obtained by taking the conditional expectation of regression equation (3*) for any given values of the three explanatory variables $X_{i1}$, $X_{i2}$, and $M_i$:

$$\begin{aligned} E(Y_i | X_{i1}, X_{i2}, M_i) = \alpha_0 + \alpha_1 X_{i1} + \alpha_2 X_{i2} + \alpha_3 X_{i1}^2 + \alpha_4 X_{i2}^2 + \alpha_5 X_{i1} X_{i2} \\ + \gamma_0 M_i + \gamma_1 M_i X_{i1} + \gamma_2 M_i X_{i2} + \gamma_3 M_i X_{i1}^2 + \gamma_4 M_i X_{i2}^2 + \gamma_5 M_i X_{i1} X_{i2} \end{aligned} \quad (3.1^*)$$

***Derivation of Model 3* from Model 3***

**Substitute** for the female dummy variable $\mathbf{F_i}$ in the PRE for Model 3 the equivalent expression $\mathbf{F_i = 1 - M_i}$:

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}^2 + \beta_4 X_{i2}^2 + \beta_5 X_{i1} X_{i2} \\ &\quad + \delta_0 F_i + \delta_1 F_i X_{i1} + \delta_2 F_i X_{i2} + \delta_3 F_i X_{i1}^2 + \delta_4 F_i X_{i2}^2 + \delta_5 F_i X_{i1} X_{i2} + u_i \end{aligned} \qquad \textbf{(3)}$$

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}^2 + \beta_4 X_{i2}^2 + \beta_5 X_{i1} X_{i2} \\ &\quad + \delta_0 (1 - M_i) + \delta_1 (1 - M_i) X_{i1} + \delta_2 (1 - M_i) X_{i2} + \delta_3 (1 - M_i) X_{i1}^2 + \delta_4 (1 - M_i) X_{i2}^2 + \delta_5 (1 - M_i) X_{i1} X_{i2} + u_i \end{aligned}$$

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}^2 + \beta_4 X_{i2}^2 + \beta_5 X_{i1} X_{i2} + \delta_0 + \delta_1 X_{i1} + \delta_2 X_{i2} + \delta_3 X_{i1}^2 + \delta_4 X_{i2}^2 + \delta_5 X_{i1} X_{i2} \\ &\quad - \delta_0 M_i - \delta_1 M_i X_{i1} - \delta_2 M_i X_{i2} - \delta_3 M_i X_{i1}^2 - \delta_4 M_i X_{i2}^2 - \delta_5 M_i X_{i1} X_{i2} + u_i \end{aligned}$$

$$\begin{aligned} Y_i &= (\beta_0 + \delta_0) + (\beta_1 + \delta_1) X_{i1} + (\beta_2 + \delta_2) X_{i2} + (\beta_3 + \delta_3) X_{i1}^2 + (\beta_4 + \delta_4) X_{i2}^2 + (\beta_5 + \delta_5) X_{i1} X_{i2} \\ &\quad - \delta_0 M_i - \delta_1 M_i X_{i1} - \delta_2 M_i X_{i2} - \delta_3 M_i X_{i1}^2 - \delta_4 M_i X_{i2}^2 - \delta_5 M_i X_{i1} X_{i2} + u_i \end{aligned}$$

$$\begin{aligned} Y_i &= \alpha_0 + \alpha_1 X_{i1} + \alpha_2 X_{i2} + \alpha_3 X_{i1}^2 + \alpha_4 X_{i2}^2 + \alpha_5 X_{i1} X_{i2} \\ &\quad + \gamma_0 M_i + \gamma_1 M_i X_{i1} + \gamma_2 M_i X_{i2} + \gamma_3 M_i X_{i1}^2 + \gamma_4 M_i X_{i2}^2 + \gamma_5 M_i X_{i1} X_{i2} + u_i \end{aligned} \qquad \textbf{(3*)}$$

where the last line comes from defining $\boldsymbol{\alpha_j = \beta_j + \delta_j}$ and $\boldsymbol{\gamma_j = -\delta_j}$ for all j = 0, 1, …, 5.

$\boldsymbol{\alpha_j}$ = the ***female*** coefficient of regressor j (j = 0, 1, …, 5)

$\boldsymbol{\gamma_j}$ **=** ***male*** coefficient of regressor j *minus* ***female*** coefficient of regressor j (j = 0, 1, …, 5)

$$\begin{aligned} E(Y_i \mid X_{i1}, X_{i2}, M_i) &= \alpha_0 + \alpha_1 X_{i1} + \alpha_2 X_{i2} + \alpha_3 X_{i1}^2 + \alpha_4 X_{i2}^2 + \alpha_5 X_{i1} X_{i2} \\ &\quad + \gamma_0 M_i + \gamma_1 M_i X_{i1} + \gamma_2 M_i X_{i2} + \gamma_3 M_i X_{i1}^2 + \gamma_4 M_i X_{i2}^2 + \gamma_5 M_i X_{i1} X_{i2} \end{aligned} \tag{3.1*}$$

- The ***male*** **population regression function**, or conditional mean function, implied by Model 3* is obtained by setting the male indicator variable $M_i = 1$ in (3.1*):

$$\begin{aligned} E(Y_i \mid X_{i1}, X_{i2}, M_i = 1) &= \alpha_0 + \alpha_1 X_{i1} + \alpha_2 X_{i2} + \alpha_3 X_{i1}^2 + \alpha_4 X_{i2}^2 + \alpha_5 X_{i1} X_{i2} \\ &\quad + \gamma_0 + \gamma_1 X_{i1} + \gamma_2 X_{i2} + \gamma_3 X_{i1}^2 + \gamma_4 X_{i2}^2 + \gamma_5 X_{i1} X_{i2} \\ &= (\alpha_0 + \gamma_0) + (\alpha_1 + \gamma_1) X_{i1} + (\alpha_2 + \gamma_2) X_{i2} \\ &\quad + (\alpha_3 + \gamma_3) X_{i1}^2 + (\alpha_4 + \gamma_4) X_{i2}^2 + (\alpha_5 + \gamma_5) X_{i1} X_{i2} \\ &= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}^2 + \beta_4 X_{i2}^2 + \beta_5 X_{i1} X_{i2} \end{aligned} \tag{3.2*}$$

where the male regression coefficients are $\beta_j = \alpha_j + \gamma_j$ for all j = 0, 1, …, 5.

- The ***female*** **population regression function**, or conditional mean function, implied by Model 3* is obtained by setting the male indicator variable $M_i = 0$ in (3.1*):

$$E(Y_i \mid X_{i1}, X_{i2}, M_i = 0) = \alpha_0 + \alpha_1 X_{i1} + \alpha_2 X_{i2} + \alpha_3 X_{i1}^2 + \alpha_4 X_{i2}^2 + \alpha_5 X_{i1} X_{i2} \tag{3.3*}$$

- The ***difference* between the *male* and *female* regression functions** – that is, the ***male-female difference* in mean Y** for given (equal) values of the explanatory variables $X_1$ and $X_2$ – is:

$$E(Y_i \mid X_{i1}, X_{i2}, M_i = 1) - E(Y_i \mid X_{i1}, X_{i2}, M_i = 0)$$
$$= \gamma_0 + \gamma_1 X_{i1} + \gamma_2 X_{i2} + \gamma_3 X_{i1}^2 + \gamma_4 X_{i2}^2 + \gamma_5 X_{i1} X_{i2}$$

- ***Interpretation* of the regression coefficients $\gamma_j$ (j = 0, 1, …, 5) in Model 3***

Each of the $\gamma_j$ coefficients in Model 3* equals a ***male*** regression coefficient *minus* the corresponding ***female*** regression coefficient: $\gamma_j = \beta_j - \alpha_j$ for all j.

$\gamma_0 = \beta_0 - \alpha_0$ = *male* intercept coefficient – *female* intercept coefficient
$\gamma_1 = \beta_1 - \alpha_1$ = *male* slope coefficient of $X_{i1}$ – *female* slope coefficient of $X_{i1}$
$\gamma_2 = \beta_2 - \alpha_2$ = *male* slope coefficient of $X_{i2}$ – *female* slope coefficient of $X_{i2}$
$\gamma_3 = \beta_3 - \alpha_3$ = *male* slope coefficient of $X_{i1}^2$ – *female* slope coefficient of $X_{i1}^2$
$\gamma_4 = \beta_4 - \alpha_4$ = *male* slope coefficient of $X_{i2}^2$ – *female* slope coefficient of $X_{i2}^2$
$\gamma_5 = \beta_5 - \alpha_5$ = *male* slope coefficient of $X_{i1}X_{i2}$ – *female* slope coefficient of $X_{i1}X_{i2}$

i.e., **$\gamma_j$ = *male*** coefficient of regressor j *minus* ***female*** coefficient of regressor j (j = 0, 1, …, 5)

- ***Recall the interpretation* of the regression coefficients $\delta_j$ (j = 0, 1, …, 5) on the female dummy interaction terms in Model 3**

  Each of the $\delta_j$ coefficients in Model 3 equals a ***female*** regression coefficient *minus* the corresponding ***male*** regression coefficient: $\delta_j = \alpha_j - \beta_j$ for all j.

  $\delta_0 = \alpha_0 - \beta_0$ = *female* intercept coefficient – *male* intercept coefficient
  $\delta_1 = \alpha_1 - \beta_1$ = *female* slope coefficient of $X_{i1}$ – *male* slope coefficient of $X_{i1}$
  $\delta_2 = \alpha_2 - \beta_2$ = *female* slope coefficient of $X_{i2}$ – *male* slope coefficient of $X_{i2}$
  $\delta_3 = \alpha_3 - \beta_3$ = *female* slope coefficient of $X_{i1}^2$ – *male* slope coefficient of $X_{i1}^2$
  $\delta_4 = \alpha_4 - \beta_4$ = *female* slope coefficient of $X_{i2}^2$ – *male* slope coefficient of $X_{i2}^2$
  $\delta_5 = \alpha_5 - \beta_5$ = *female* slope coefficient of $X_{i1}X_{i2}$ – *male* slope coefficient of $X_{i1}X_{i2}$

  i.e., **$\delta_j$ = *female*** coefficient of regressor j *minus* ***male*** coefficient of regressor j (j = 0, 1, …, 5)

- ***<u>RESULT</u>:* The regression coefficients $\delta_j$ (j = 0, 1, …, 5) on the *female* dummy interaction terms in Model 3 equal *the negative of* the regression coefficients $\gamma_j$ (j = 0, 1, …, 5) on the *male* dummy interaction terms in Model 3***:

  $\boldsymbol{\delta_j = -\gamma_j}$ for all j = 0, 1, …, 5

***Compare Model 3 and Model 3*:*** they are **observationally equivalent**

***Model 3:*** **a full-interaction regression equation** in the ***female*** **dummy variable $F_i$**

$$\begin{aligned} Y_i = {} & \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}^2 + \beta_4 X_{i2}^2 + \beta_5 X_{i1} X_{i2} \\ & + \delta_0 F_i + \delta_1 F_i X_{i1} + \delta_2 F_i X_{i2} + \delta_3 F_i X_{i1}^2 + \delta_4 F_i X_{i2}^2 + \delta_5 F_i X_{i1} X_{i2} + u_i \end{aligned} \qquad \textbf{(3)}$$

where

$\beta_j$ = male regression coefficients

$\alpha_j = \beta_j + \delta_j$ = female regression coefficients

$\delta_j = \alpha_j - \beta_j$ = female-male coefficient differences

***Model 3*:*** **an alternative full-interaction regression equation** in the ***male*** **dummy variable $M_i$**

$$\begin{aligned} Y_i = {} & \alpha_0 + \alpha_1 X_{i1} + \alpha_2 X_{i2} + \alpha_3 X_{i1}^2 + \alpha_4 X_{i2}^2 + \alpha_5 X_{i1} X_{i2} \\ & + \gamma_0 M_i + \gamma_1 M_i X_{i1} + \gamma_2 M_i X_{i2} + \gamma_3 M_i X_{i1}^2 + \gamma_4 M_i X_{i2}^2 + \gamma_5 M_i X_{i1} X_{i2} + u_i \end{aligned} \qquad \textbf{(3*)}$$

where

$\alpha_j$ = female regression coefficients

$\beta_j = \alpha_j + \gamma_j$ = male regression coefficients

$\gamma_j = \beta_j - \alpha_j$ = male-female coefficient differences

**<u>Relationship between OLS coefficient estimates of Models 3 and 3*</u>:** the OLS coefficient estimates from Models 3 and 3* are identical.

$\hat{\beta}_j$ from Model 3 $=$ $\hat{\alpha}_j + \hat{\gamma}_j$ from Model 3*

$\hat{\alpha}_j = \hat{\beta}_j + \hat{\delta}_j$ from Model 3 $=$ $\hat{\alpha}_j$ from Model 3*

$\hat{\delta}_j = \hat{\alpha}_j - \hat{\beta}_j$ from Model 3 $=$ $-\hat{\gamma}_j = -(\hat{\beta}_j - \hat{\alpha}_j)$ from Model 3*