ECON 351* -- Introduction to NOTE 21

Introduction to Dummy Variable Regressors

1. An Example of Dummy Variable Regressors

- A model of North American car prices given by the PRE

$$price_i = \beta_1 + \beta_2 wgt_i + \beta_3 wgt_i^2 + \beta_4 mpg_i + \beta_5 frn_i + \beta_6 frn_i wgt_i + \beta_7 frn_i wgt_i^2 + u_i \quad \textbf{(3)}$$

where

$price_i$ = the price of the i-th car (in US dollars);
wgt_i = the weight of the i-th car (in pounds);
mpg_i = the fuel efficiency of the i-th car (in miles per gallon);
frn_i = 1 if the i-th car is foreign, = 0 if the i-th car is domestic;
N = 74 = the number of observations in the estimation sample.

- The **regressor frn_i** is a ***binary*** **variable** called an ***indicator*** **or** ***dummy*** **variable.**

By definition, the ***binary*** **variable frn_i** takes only ***two*** **values**:

frn_i = 1 if the i-th car is a ***foreign*** car, meaning it is manufactured *outside* North America;
frn_i = 0 if the i-th car is a ***domestic*** car, meaning it is manufactured *inside* North America.

Because by definition **frn_i = 1** for foreign cars, it is called a foreign-car indicator or dummy variable.

- The key to ***interpreting*** **regression equation (3)** is to recognize that it in fact includes ***two*** **distinct regression models** for car prices -- one for domestic cars, the other for foreign cars.

- The **regression equation for *domestic* cars**

Set dummy variable **$frn_i = 0$** in equation (3):

$$price_i = \beta_1 + \beta_2 wgt_i + \beta_3 wgt_i^2 + \beta_4 mpg_i + \beta_5 frn_i + \beta_6 frn_i wgt_i + \beta_7 frn_i wgt_i^2 + u_i \quad \textbf{(3)}$$

$$\begin{aligned} price_i &= \beta_1 + \beta_2 wgt_i + \beta_3 wgt_i^2 + \beta_4 mpg_i + \beta_5 frn_i + \beta_6 frn_i wgt_i + \beta_7 frn_i wgt_i^2 + u_i \\ &= \beta_1 + \beta_2 wgt_i + \beta_3 wgt_i^2 + \beta_4 mpg_i + \beta_5 0 + \beta_6 (0) wgt_i + \beta_7 (0) wgt_i^2 + u_i \\ &= \beta_1 + \beta_2 wgt_i + \beta_3 wgt_i^2 + \beta_4 mpg_i + u_i \qquad \textbf{(3d)} \end{aligned}$$

- The **regression equation for *foreign* cars**

Set dummy variable **$frn_i = 1$** in equation (3):

$$price_i = \beta_1 + \beta_2 wgt_i + \beta_3 wgt_i^2 + \beta_4 mpg_i + \beta_5 frn_i + \beta_6 frn_i wgt_i + \beta_7 frn_i wgt_i^2 + u_i \quad \textbf{(3)}$$

$$\begin{aligned} price_i &= \beta_1 + \beta_2 wgt_i + \beta_3 wgt_i^2 + \beta_4 mpg_i + \beta_5 frn_i + \beta_6 frn_i wgt_i + \beta_7 frn_i wgt_i^2 + u_i \\ &= \beta_1 + \beta_2 wgt_i + \beta_3 wgt_i^2 + \beta_4 mpg_i + \beta_5 1 + \beta_6 (1) wgt_i + \beta_7 (1) wgt_i^2 + u_i \\ &= \beta_1 + \beta_2 wgt_i + \beta_3 wgt_i^2 + \beta_4 mpg_i + \beta_5 + \beta_6 wgt_i + \beta_7 wgt_i^2 + u_i \\ &= (\beta_1 + \beta_5) + (\beta_2 + \beta_6) wgt_i + (\beta_3 + \beta_7) wgt_i^2 + \beta_4 mpg_i + u_i \qquad \textbf{(3f)} \end{aligned}$$

Note that in the foreign-car price equation (3f),

- **foreign-car *intercept* coefficient** = $\beta_1 + \beta_5$
- **foreign-car *slope* coefficient on *wgt_i*** = $\beta_2 + \beta_6$
- **foreign-car *slope* coefficient on *wgt_i-squared*** = $\beta_3 + \beta_7$

- **Compare *foreign-car equation* (3f) with *domestic-car* equation (3d):**

$$\text{price}_i = \beta_1 + \beta_2 \text{wgt}_i + \beta_3 \text{wgt}_i^2 + \beta_4 \text{mpg}_i + u_i \qquad \textbf{(3d)}$$

$$\text{price}_i = (\beta_1 + \beta_5) + (\beta_2 + \beta_6)\text{wgt}_i + (\beta_3 + \beta_7)\text{wgt}_i^2 + \beta_4 \text{mpg}_i + u_i \qquad \textbf{(3f)}$$

Question: How are the regression coefficients β_5, β_6 and β_7 in regression (3) interpreted?

$$\text{price}_i = \beta_1 + \beta_2 \text{wgt}_i + \beta_3 \text{wgt}_i^2 + \beta_4 \text{mpg}_i + \beta_5 \text{frn}_i + \beta_6 \text{frn}_i \text{wgt}_i + \beta_7 \text{frn}_i \text{wgt}_i^2 + u_i \qquad \textbf{(3)}$$

Answer: By inspection and comparison of the ***domestic-car* equation (3d)** and the ***foreign-car* equation (3f)**, we see that

β_5 = foreign intercept ($\beta_1 + \beta_5$) – domestic intercept (β_1)

β_6 = foreign coefficient of wgt_i ($\beta_2 + \beta_6$) – domestic coefficient of wgt_i (β_2)

β_7 = foreign coefficient of wgt_i^2 ($\beta_3 + \beta_7$) – domestic coefficient of wgt_i^2 (β_3)

2. How Dummy Variable Regressors Enter Regression Models

- ***Indicator* (*dummy*) variables** enter as regressors in linear regression models in one of two basic ways.

1. **As Additive Regressors: Differences in Intercepts**

   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.

2. **As Multiplicative Regressors: Dummy Variable Interaction Terms**

   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 regressor (either a continuous variable or another dummy variable). They allow for **different *slope* coefficients** across identifiable subsets of observations in the population.

3. Four Different Models of North American Car Prices

- To illustrate the use of indicator (dummy) variables as regressors in linear regression models, consider the following four linear regression models for North American car prices.

Model 1: Contains no dummy variable regressors. Allows for ***no*** **coefficient differences** between ***foreign*** **and** ***domestic*** **cars**.

$$\text{price}_i = \beta_1 + \beta_2 \text{wgt}_i + \beta_3 \text{wgt}_i^2 + \beta_4 \text{mpg}_i + u_i \qquad \textbf{(1)}$$

Model 2: Allows for **different foreign-car and domestic-car** ***intercepts*** by introducing the foreign-car indicator variable frn_i as an additional ***additive*** **regressor** in Model 1.

$$\text{price}_i = \beta_1 + \beta_2 \text{wgt}_i + \beta_3 \text{wgt}_i^2 + \beta_4 \text{mpg}_i + \delta_1 \text{frn}_i + u_i \qquad \textbf{(2)}$$

Model 3: Allows for (1) **different foreign-car and domestic-car** ***intercepts*** and (2) different **foreign-car and domestic-car** ***slope coefficients*** on the regressors wgt_i and wgt_i^2. Introduces the **foreign-car interaction terms** frn_iwgt_i and $\text{frn}_i\text{wgt}_i^2$ as additional ***multiplicative*** **regressors** in Model 2.

$$\text{price}_i = \beta_1 + \beta_2 \text{wgt}_i + \beta_3 \text{wgt}_i^2 + \beta_4 \text{mpg}_i + \delta_1 \text{frn}_i + \delta_2 \text{frn}_i \text{wgt}_i + \delta_3 \text{frn}_i \text{wgt}_i^2 + u_i \qquad \textbf{(3)}$$

Model 4: Allows ***all*** **regression coefficients** -- both ***intercept*** **and** ***slope*** **coefficients** -- to differ between foreign and domestic cars. It allows for (1) **different foreign-car and domestic-car** ***intercepts*** and (2) different **foreign-car and domestic-car** ***slope coefficients*** on ***all three*** **regressors** in Model 1, namely wgt_i, wgt_i^2, and mpg_i. Introduces the foreign-car interaction term frn_impg_i as an additional ***multiplicative*** **regressor** in Model 3.

$$\begin{aligned}\text{price}_i = {} & \beta_1 + \beta_2 \text{wgt}_i + \beta_3 \text{wgt}_i^2 + \beta_4 \text{mpg}_i \\ & + \delta_1 \text{frn}_i + \delta_2 \text{frn}_i \text{wgt}_i + \delta_3 \text{frn}_i \text{wgt}_i^2 + \delta_4 \text{frn}_i \text{mpg}_i + u_i \qquad \textbf{(4)}\end{aligned}$$

4. Interpreting Model 4: A Full-Interaction Regression Model

Model 4

$$\begin{aligned} \text{price}_i = {} & \beta_1 + \beta_2 \text{wgt}_i + \beta_3 \text{wgt}_i^2 + \beta_4 \text{mpg}_i \\ & + \delta_1 \text{frn}_i + \delta_2 \text{frn}_i \text{wgt}_i + \delta_3 \text{frn}_i \text{wgt}_i^2 + \delta_4 \text{frn}_i \text{mpg}_i + u_i \end{aligned} \qquad \textbf{(4)}$$

- The ***population regression function*** **for Model 4** is obtained by taking the conditional expectation of regression equation (4) for any given values of the three explanatory variables wgt_i, mpg_i and frn_i:

$$\begin{aligned} E(\text{price}_i \mid \text{wgt}_i, \text{mpg}_i, \text{frn}_i) = {} & \beta_1 + \beta_2 \text{wgt}_i + \beta_3 \text{wgt}_i^2 + \beta_4 \text{mpg}_i \\ & + \delta_1 \text{frn}_i + \delta_2 \text{frn}_i \text{wgt}_i + \delta_3 \text{frn}_i \text{wgt}_i^2 + \delta_4 \text{frn}_i \text{mpg}_i \end{aligned} \qquad (4.1)$$

- The ***domestic-car*** **regression equation** and ***domestic-car*** **regression function** are obtained by setting the foreign-car indicator variable $\mathbf{frn_i = 0}$ in (4) and (4.1):

$$\text{price}_i = \beta_1 + \beta_2 \text{wgt}_i + \beta_3 \text{wgt}_i^2 + \beta_4 \text{mpg}_i + u_i \qquad \textbf{(4d)}$$

$$E(\text{price}_i \mid \text{wgt}_i, \text{mpg}_i, \text{frn}_i = 0) = \beta_1 + \beta_2 \text{wgt}_i + \beta_3 \text{wgt}_i^2 + \beta_4 \text{mpg}_i \qquad (4.2)$$

- The ***domestic-car*** **regression coefficients** are β_j for all j = 1, …, 4:

  domestic-car intercept coefficient $= \beta_1$
  domestic-car slope coefficient of wgt_i $= \beta_2$
  domestic-car slope coefficient of wgt_i^2 $= \beta_3$
  domestic-car slope coefficient of mpg_i $= \beta_4$

$$\text{price}_i = \beta_1 + \beta_2 \text{wgt}_i + \beta_3 \text{wgt}_i^2 + \beta_4 \text{mpg}_i + \delta_1 \text{frn}_i + \delta_2 \text{frn}_i \text{wgt}_i + \delta_3 \text{frn}_i \text{wgt}_i^2 + \delta_4 \text{frn}_i \text{mpg}_i + u_i \quad \textbf{(4)}$$

$$E(\text{price}_i \mid \text{wgt}_i, \text{mpg}_i, \text{frn}_i) = \beta_1 + \beta_2 \text{wgt}_i + \beta_3 \text{wgt}_i^2 + \beta_4 \text{mpg}_i + \delta_1 \text{frn}_i + \delta_2 \text{frn}_i \text{wgt}_i + \delta_3 \text{frn}_i \text{wgt}_i^2 + \delta_4 \text{frn}_i \text{mpg}_i \quad (4.1)$$

- The ***foreign-car*** **regression equation** and ***foreign-car*** **regression function** are obtained by setting the foreign-car indicator variable $\mathbf{frn_i = 1}$ in (4) and (4.1):

$$\begin{aligned} \text{price}_i &= \beta_1 + \beta_2 \text{wgt}_i + \beta_3 \text{wgt}_i^2 + \beta_4 \text{mpg}_i + \delta_1 + \delta_2 \text{wgt}_i + \delta_3 \text{wgt}_i^2 + \delta_4 \text{mpg}_i + u_i \\ &= (\beta_1 + \delta_1) + (\beta_2 + \delta_2)\text{wgt}_i + (\beta_3 + \delta_3)\text{wgt}_i^2 + (\beta_4 + \delta_4)\text{mpg}_i + u_i \\ &= \alpha_1 + \alpha_2 \text{wgt}_i + \alpha_3 \text{wgt}_i^2 + \alpha_4 \text{mpg}_i + u_i \qquad \textbf{(4f)} \end{aligned}$$

$$\begin{aligned} E(\text{price}_i \mid \text{wgt}_i, \text{mpg}_i, \text{frn}_i = 1) \\ &= \beta_1 + \beta_2 \text{wgt}_i + \beta_3 \text{wgt}_i^2 + \beta_4 \text{mpg}_i + \delta_1 + \delta_2 \text{wgt}_i + \delta_3 \text{wgt}_i^2 + \delta_4 \text{mpg}_i \\ &= (\beta_1 + \delta_1) + (\beta_2 + \delta_2)\text{wgt}_i + (\beta_3 + \delta_3)\text{wgt}_i^2 + (\beta_4 + \delta_4)\text{mpg}_i \\ &= \alpha_1 + \alpha_2 \text{wgt}_i + \alpha_3 \text{wgt}_i^2 + \alpha_4 \text{mpg}_i \qquad (4.3) \end{aligned}$$

- The ***foreign-car*** **regression coefficients** are $\alpha_j = \beta_j + \delta_j$ for all j = 1, …, 4:

  foreign-car intercept coefficient $= \alpha_1 = \beta_1 + \delta_1$
  foreign-car slope coefficient of wgt_i $= \alpha_2 = \beta_2 + \delta_2$
  foreign-car slope coefficient of wgt_i^2 $= \alpha_3 = \beta_3 + \delta_3$
  foreign-car slope coefficient of mpg_i $= \alpha_4 = \beta_4 + \delta_4$

- Solving the equations $\alpha_j = \beta_j + \delta_j$ for δ_j yields the result $\boldsymbol{\delta_j = \alpha_j - \beta_j}$ for j = 1, …, 4.

  This gives us the interpretation of the δ_j coefficients in Model 4.

- ***Interpretation* of the regression coefficients δ_j (j = 1, …, 4) in Model 4**

$$\text{price}_i = \beta_1 + \beta_2 \text{wgt}_i + \beta_3 \text{wgt}_i^2 + \beta_4 \text{mpg}_i + \delta_1 \text{frn}_i + \delta_2 \text{frn}_i \text{wgt}_i + \delta_3 \text{frn}_i \text{wgt}_i^2 + \delta_4 \text{frn}_i \text{mpg}_i + u_i \qquad \mathbf{(4)}$$

$$E(\text{price}_i \mid \text{wgt}_i, \text{mpg}_i, \text{frn}_i) = \beta_1 + \beta_2 \text{wgt}_i + \beta_3 \text{wgt}_i^2 + \beta_4 \text{mpg}_i + \delta_1 \text{frn}_i + \delta_2 \text{frn}_i \text{wgt}_i + \delta_3 \text{frn}_i \text{wgt}_i^2 + \delta_4 \text{frn}_i \text{mpg}_i \qquad (4.1)$$

Each of the δ_j coefficients in Model 4 equals a ***foreign-car*** regression coefficient *minus* the corresponding ***domestic-car*** regression coefficient: $\delta_j = \alpha_j - \beta_j$ for all j.

$\delta_1 = \alpha_1 - \beta_1$
= *foreign* intercept coefficient − *domestic* intercept coefficient

$\delta_2 = \alpha_2 - \beta_2$
= *foreign* slope coefficient of wgt_i − *domestic* slope coefficient of wgt_i

$\delta_3 = \alpha_3 - \beta_3$
= *foreign* slope coefficient of wgt_i^2 − *domestic* slope coefficient of wgt_i^2

$\delta_4 = \alpha_4 - \beta_4$
= *foreign* slope coefficient of mpg_i − *domestic* slope coefficient of mpg_i

- The ***difference*** **between the** ***foreign-car*** **regression function and the** ***domestic-car*** **regression function** is the ***foreign-domestic car difference*** **in mean car prices** for given ***equal*** **values** of the explanatory variables wgt_i and mpg_i.

$$E(price_i \mid wgt_i, mpg_i, frn_i = 1)$$
$$= \beta_1 + \beta_2 wgt_i + \beta_3 wgt_i^2 + \beta_4 mpg_i + \delta_1 + \delta_2 wgt_i + \delta_3 wgt_i^2 + \delta_4 mpg_i \quad (4.3)$$

$$E(price_i \mid wgt_i, mpg_i, frn_i = 0) = \beta_1 + \beta_2 wgt_i + \beta_3 wgt_i^2 + \beta_4 mpg_i \quad (4.2)$$

Subtract equation (4.2) for domestic cars from equation (4.3) for foreign cars:

$$E(price_i \mid wgt_i, mpg_i, frn_i = 1) - E(price_i \mid wgt_i, mpg_i, frn_i = 0)$$
$$= \beta_1 + \beta_2 wgt_i + \beta_3 wgt_i^2 + \beta_4 mpg_i + \delta_1 + \delta_2 wgt_i + \delta_3 wgt_i^2 + \delta_4 mpg_i$$
$$- \beta_1 - \beta_2 wgt_i - \beta_3 wgt_i^2 - \beta_4 mpg_i$$
$$= \delta_1 + \delta_2 wgt_i + \delta_3 wgt_i^2 + \delta_4 mpg_i$$

Result:

$$E(price_i \mid wgt_i, mpg_i, frn_i = 1) - E(price_i \mid wgt_i, mpg_i, frn_i = 0)$$
$$= \delta_1 + \delta_2 wgt_i + \delta_3 wgt_i^2 + \delta_4 mpg_i$$

Interpretation:

- The foreign-domestic difference in the conditional mean value of car price for given values wgt_i and mpg_i of the explanatory variables wgt and mpg is **a** ***function*** **of** $\mathbf{wgt_i}$ **and** $\mathbf{mpg_i}$. It is ***not*** **a** ***constant***, but instead depends on the values of the explanatory variables wgt and mpg.

- The conditional foreign-domestic mean car price difference addresses the following question: **What is the** ***foreign-domestic*** **difference in mean car price for** ***identical*** **(*equal*) values of the explanatory variables** ***wgt*** **and** ***mpg*****? What is the** ***mean price difference*** **between foreign and domestic cars of the** ***same*** **size (wgt) and fuel efficiency (mpg)?**

5. An Alternative Estimating Equation for Model 4

The regression equation for Model 4 can be written in an alternative but equivalent way.

- **Define a Domestic Car Indicator Variable**

  Define an **indicator or dummy variable** for ***domestic cars*** named $\boldsymbol{dom_i}$:

  $\mathbf{dom_i = 1}$ if the i-th car is a ***domestic*** car, meaning it is manufactured *inside* North America;
  $\mathbf{dom_i = 0}$ if the i-th car is a ***foreign*** car, meaning it is manufactured *outside* North America.

  By definition, the domestic car indicator variable $\boldsymbol{dom_i}$ is related to the foreign car indicator variable $\boldsymbol{frn_i}$ as follows:

  $\mathbf{dom_i = 1 - frn_i}$ for all i

  $\mathbf{dom_i + frn_i = 1}$ so that $\mathbf{frn_i = 1 - dom_i}$ for all i

  - For domestic cars: $\mathbf{frn_i = 0}$ and $\mathbf{dom_i = 1}$
  - For foreign cars: $\mathbf{frn_i = 1}$ and $\mathbf{dom_i = 0}$

- **One Estimating Equation for Model 4**

  The estimating equation for Model 4 we have used so far includes a full set of **interaction terms** in the **foreign car indicator variable** $\boldsymbol{frn_i}$:

$$\begin{aligned}\text{price}_i = {} & \beta_1 + \beta_2 \text{wgt}_i + \beta_3 \text{wgt}_i^2 + \beta_4 \text{mpg}_i \\ & + \delta_1 \text{frn}_i + \delta_2 \text{frn}_i \text{wgt}_i + \delta_3 \text{frn}_i \text{wgt}_i^2 + \delta_4 \text{frn}_i \text{mpg}_i + u_i \qquad \textbf{(4A)}\end{aligned}$$

  The car type whose dummy variable is excluded from equation (4A) is domestic cars; ***domestic cars*** therefore constitute the **base group** for car type in equation (4A).

- **Derivation of a Second Estimating Equation for Model 4**

In equation (4A), substitute for the foreign indicator variable $\mathbf{frn_i}$ the equivalent expression $\mathbf{1 - dom_i}$; i.e., set $\mathbf{frn_i = 1 - dom_i}$ in equation (4A).

$$\begin{aligned}\text{price}_i = {} & \beta_1 + \beta_2 \text{wgt}_i + \beta_3 \text{wgt}_i^2 + \beta_4 \text{mpg}_i \\ & + \delta_1 \text{frn}_i + \delta_2 \text{frn}_i \text{wgt}_i + \delta_3 \text{frn}_i \text{wgt}_i^2 + \delta_4 \text{frn}_i \text{mpg}_i + u_i \end{aligned} \qquad \textbf{(4A)}$$

1. $$\begin{aligned}\text{price}_i = {} & \beta_1 + \beta_2 \text{wgt}_i + \beta_3 \text{wgt}_i^2 + \beta_4 \text{mpg}_i + \delta_1 (1 - \text{dom}_i) \\ & + \delta_2 (1 - \text{dom}_i) \text{wgt}_i + \delta_3 (1 - \text{dom}_i) \text{wgt}_i^2 + \delta_4 (1 - \text{dom}_i) \text{mpg}_i + u_i \end{aligned}$$

2. $$\begin{aligned}\text{price}_i = {} & \beta_1 + \beta_2 \text{wgt}_i + \beta_3 \text{wgt}_i^2 + \beta_4 \text{mpg}_i + \delta_1 + \delta_2 \text{wgt}_i + \delta_3 \text{wgt}_i^2 + \delta_4 \text{mpg}_i \\ & - \delta_1 \text{dom}_i - \delta_2 \text{dom}_i \text{wgt}_i - \delta_3 \text{dom}_i \text{wgt}_i^2 - \delta_4 \text{dom}_i \text{mpg}_i + u_i \end{aligned}$$

3. $$\begin{aligned}\text{price}_i = {} & (\beta_1 + \delta_1) + (\beta_2 + \delta_2) \text{wgt}_i + (\beta_3 + \delta_3) \text{wgt}_i^2 + (\beta_4 + \delta_4) \text{mpg}_i \\ & - \delta_1 \text{dom}_i - \delta_2 \text{dom}_i \text{wgt}_i - \delta_3 \text{dom}_i \text{wgt}_i^2 - \delta_4 \text{dom}_i \text{mpg}_i + u_i \end{aligned}$$

In the foreign-car regression equation (4f), we previously defined the coefficients $\beta_j + \delta_j$ as α_j (j = 1, 2, 3, 4), the foreign-car regression coefficients. In the above regression equation set $\beta_j + \delta_j = \alpha_j$ for j = 1, 2, 3, 4:

$$\begin{aligned}\text{price}_i = {} & \alpha_1 + \alpha_2 \text{wgt}_i + \alpha_3 \text{wgt}_i^2 + \alpha_4 \text{mpg}_i \\ & - \delta_1 \text{dom}_i - \delta_2 \text{dom}_i \text{wgt}_i - \delta_3 \text{dom}_i \text{wgt}_i^2 - \delta_4 \text{dom}_i \text{mpg}_i + u_i \end{aligned}$$

Finally, replace the $-\delta_j$ coefficients with γ_j for j = 1, 2, 3, 4:

$$\begin{aligned}\text{price}_i = {} & \alpha_1 + \alpha_2 \text{wgt}_i + \alpha_3 \text{wgt}_i^2 + \alpha_4 \text{mpg}_i \\ & + \gamma_1 \text{dom}_i + \gamma_2 \text{dom}_i \text{wgt}_i + \gamma_3 \text{dom}_i \text{wgt}_i^2 + \gamma_4 \text{dom}_i \text{mpg}_i + u_i \end{aligned} \qquad \textbf{(4B)}$$

Regression equation (4B) is a second estimating equation for Model 4; it is ***observationally equivalent*** to regression equation (4A). ***Foreign cars*** constitute the **base group** for car type in equation (4B).

- **Interpretation of Second Estimating Equation (4B) for Model 4**

  Equation (4B) and its implied regression function are:

$$\text{price}_i = \alpha_1 + \alpha_2 \text{wgt}_i + \alpha_3 \text{wgt}_i^2 + \alpha_4 \text{mpg}_i + \gamma_1 \text{dom}_i + \gamma_2 \text{dom}_i \text{wgt}_i + \gamma_3 \text{dom}_i \text{wgt}_i^2 + \gamma_4 \text{dom}_i \text{mpg}_i + u_i \qquad \textbf{(4B)}$$

$$E(\text{price}_i \mid \text{wgt}_i, \text{mpg}_i, \text{dom}_i) = \alpha_1 + \alpha_2 \text{wgt}_i + \alpha_3 \text{wgt}_i^2 + \alpha_4 \text{mpg}_i + \gamma_1 \text{dom}_i + \gamma_2 \text{dom}_i \text{wgt}_i + \gamma_3 \text{dom}_i \text{wgt}_i^2 + \gamma_4 \text{dom}_i \text{mpg}_i \qquad (4B.1)$$

- The ***foreign-car*** **regression equation** and ***foreign-car*** **regression function** are obtained by setting the domestic-car indicator variable **$\text{dom}_i = 0$** in (4B) and (4B.1):

$$\text{price}_i = \alpha_1 + \alpha_2 \text{wgt}_i + \alpha_3 \text{wgt}_i^2 + \alpha_4 \text{mpg}_i + u_i \qquad \textbf{(4f)}$$

$$E(\text{price}_i \mid \text{wgt}_i, \text{mpg}_i, \text{dom}_i = 0) = \alpha_1 + \alpha_2 \text{wgt}_i + \alpha_3 \text{wgt}_i^2 + \alpha_4 \text{mpg}_i$$

- The ***domestic-car*** **regression equation** and ***domestic-car*** **regression function** are obtained by setting the domestic-car indicator variable **$\text{dom}_i = 1$** in (4B) and (4B.1):

$$\begin{aligned}\text{price}_i &= \alpha_1 + \alpha_2 \text{wgt}_i + \alpha_3 \text{wgt}_i^2 + \alpha_4 \text{mpg}_i + \gamma_1 + \gamma_2 \text{wgt}_i + \gamma_3 \text{wgt}_i^2 + \gamma_4 \text{mpg}_i + u_i \\ &= (\alpha_1 + \gamma_1) + (\alpha_2 + \gamma_2)\text{wgt}_i + (\alpha_3 + \gamma_3)\text{wgt}_i^2 + (\alpha_4 + \gamma_4)\text{mpg}_i + u_i \\ &= \beta_1 + \beta_2 \text{wgt}_i + \beta_3 \text{wgt}_i^2 + \beta_4 \text{mpg}_i + u_i \qquad \textbf{(4d)}\end{aligned}$$

$$\begin{aligned}E(\text{price}_i \mid \text{wgt}_i, \text{mpg}_i, \text{dom}_i = 1) &= \alpha_1 + \alpha_2 \text{wgt}_i + \alpha_3 \text{wgt}_i^2 + \alpha_4 \text{mpg}_i + \gamma_1 + \gamma_2 \text{wgt}_i + \gamma_3 \text{wgt}_i^2 + \gamma_4 \text{mpg}_i \\ &= (\alpha_1 + \gamma_1) + (\alpha_2 + \gamma_2)\text{wgt}_i + (\alpha_3 + \gamma_3)\text{wgt}_i^2 + (\alpha_4 + \gamma_4)\text{mpg}_i \\ &= \beta_1 + \beta_2 \text{wgt}_i + \beta_3 \text{wgt}_i^2 + \beta_4 \text{mpg}_i\end{aligned}$$

  where $\beta_j = \alpha_j + \gamma_j$ for j = 1, 2, 3, 4 are the domestic-car regression coefficients.

- **Compare Estimating Equations (4A) and (4B) for Model 4**

Equation (4A):

$$\text{price}_i = \beta_1 + \beta_2 \text{wgt}_i + \beta_3 \text{wgt}_i^2 + \beta_4 \text{mpg}_i + \delta_1 \text{frn}_i + \delta_2 \text{frn}_i \text{wgt}_i + \delta_3 \text{frn}_i \text{wgt}_i^2 + \delta_4 \text{frn}_i \text{mpg}_i + u_i \quad \textbf{(4A)}$$

where:

- the domestic-car regression coefficients are β_j (j = 1, 2, 3, 4)
- the foreign-car regression coefficients are $\alpha_j = \beta_j + \delta_j$ (j = 1, 2, 3, 4)
- implied expressions for β_j are: $\beta_j = \alpha_j - \delta_j$ (j = 1, 2, 3, 4)

Equation (4B):

$$\text{price}_i = \alpha_1 + \alpha_2 \text{wgt}_i + \alpha_3 \text{wgt}_i^2 + \alpha_4 \text{mpg}_i + \gamma_1 \text{dom}_i + \gamma_2 \text{dom}_i \text{wgt}_i + \gamma_3 \text{dom}_i \text{wgt}_i^2 + \gamma_4 \text{dom}_i \text{mpg}_i + u_i \quad \textbf{(4B)}$$

where:

- the domestic-car regression coefficients are $\beta_j = \alpha_j + \gamma_j$ (j = 1, 2, 3, 4)
- the foreign-car regression coefficients are α_j (j = 1, 2, 3, 4)
- implied expressions for α_j are: $\alpha_j = \beta_j - \gamma_j$ (j = 1, 2, 3, 4)

Compare expressions for *foreign-car* coefficients α_j from equations (4A) and (4B):

$\alpha_j = \beta_j + \delta_j$ in (4A) and $\alpha_j = \beta_j - \gamma_j$ in (4B) implies that $\delta_j = -\gamma_j$

Compare expressions for *domestic-car* coefficients β_j from equations (4A) and (4B):

$\beta_j = \alpha_j - \delta_j$ in (4A) and $\beta_j = \alpha_j + \gamma_j$ in (4B) implies that $\gamma_j = -\delta_j$

<u>Results</u>: Equations (4A) and (4B) are observationally equivalent regression equations.

$$\text{price}_i = \beta_1 + \beta_2 \text{wgt}_i + \beta_3 \text{wgt}_i^2 + \beta_4 \text{mpg}_i + \delta_1 \text{frn}_i + \delta_2 \text{frn}_i \text{wgt}_i + \delta_3 \text{frn}_i \text{wgt}_i^2 + \delta_4 \text{frn}_i \text{mpg}_i + u_i \qquad \textbf{(4A)}$$

$$\text{price}_i = \alpha_1 + \alpha_2 \text{wgt}_i + \alpha_3 \text{wgt}_i^2 + \alpha_4 \text{mpg}_i + \gamma_1 \text{dom}_i + \gamma_2 \text{dom}_i \text{wgt}_i + \gamma_3 \text{dom}_i \text{wgt}_i^2 + \gamma_4 \text{dom}_i \text{mpg}_i + u_i \qquad \textbf{(4B)}$$

1. The δ_j coefficients in (4A) and the γ_j coefficients in (4B) are **equal in magnitude** but **opposite in sign**.

δ_j coefficients = foreign-domestic coefficient differences $= \alpha_j - \beta_j$

γ_j coefficients = domestic-foreign coefficient differences $= \beta_j - \alpha_j$

2. Equations (4A) and (4B) yield ***identical*** **estimates** of the ***foreign-car*** coefficients α_j and the ***domestic-car*** coefficients β_j.

3. OLS estimation of equations (4A) and (4B) yields identical values of:

RSS = the residual sum-of-squares

ESS = the explained sum-of-squares

R^2 = the ordinary R-squared

$\bar{R}^2$ = the adjusted R-squared

$\hat{\sigma}^2$ = the estimator of the error variance σ^2

$\text{ANOVA} - F_0$ = the ANOVA F-statistic

Example: OLS estimates of Equations (4A) and (4B)

```
. * Equation (4A)
. regress price wgt wgtsq mpg frn frnwgt frnwgtsq frnmpg

      Source |       SS       df       MS              Number of obs =      74
-------------+------------------------------           F(  7,    66) =   18.65
       Model |   421816157     7  60259451.1           Prob > F      =  0.0000
    Residual |   213249239    66  3231049.07           R-squared     =  0.6642
-------------+------------------------------           Adj R-squared =  0.6286
       Total |   635065396    73  8699525.97           Root MSE      =  1797.5

------------------------------------------------------------------------------
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         wgt |  -8.813044   3.023761    -2.91   0.005    -14.85018    -2.77591
       wgtsq |   .0018982   .0004203     4.52   0.000      .001059    .0027374
         mpg |   78.07451   115.5373     0.68   0.502    -152.6031    308.7521
         frn |  -2474.355   14701.79    -0.17   0.867    -31827.42    26878.71
      frnwgt |   2.144557    10.2767     0.21   0.835    -18.37353    22.66265
    frnwgtsq |   .0004253   .0019601     0.22   0.829    -.0034882    .0043388
      frnmpg |  -100.1518   141.2379    -0.71   0.481    -382.1422    181.8387
       _cons |   11972.05   7197.498     1.66   0.101    -2398.214    26342.32
------------------------------------------------------------------------------
```

```
. * Equation (4B)
. regress price wgt wgtsq mpg dom domwgt domwgtsq dommpg

      Source |       SS       df       MS              Number of obs =      74
-------------+------------------------------           F(  7,    66) =   18.65
       Model |   421816157     7  60259451.1           Prob > F      =  0.0000
    Residual |   213249239    66  3231049.07           R-squared     =  0.6642
-------------+------------------------------           Adj R-squared =  0.6286
       Total |   635065396    73  8699525.97           Root MSE      =  1797.5

------------------------------------------------------------------------------
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         wgt |  -6.668487   9.821781    -0.68   0.500    -26.27831    12.94133
       wgtsq |   .0023235   .0019145     1.21   0.229     -.001499    .0061459
         mpg |  -22.07724   81.23592    -0.27   0.787      -184.27    140.1155
         dom |   2474.355   14701.79     0.17   0.867    -26878.71    31827.42
      domwgt |  -2.144557    10.2767    -0.21   0.835    -22.66265    18.37353
    domwgtsq |  -.0004253   .0019601    -0.22   0.829    -.0043388    .0034882
      dommpg |   100.1518   141.2379     0.71   0.481    -181.8387    382.1422
       _cons |   9497.699   12819.46     0.74   0.461    -16097.18    35092.58
------------------------------------------------------------------------------
```

Computing foreign-car coefficient estimates from Equation (4A)

```
. * Following Equation (4A)

. lincom _b[_cons] + _b[frn]

 ( 1)  frn + _cons = 0.0

------------------------------------------------------------------------------
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   9497.699   12819.46     0.74   0.461    -16097.18    35092.58
------------------------------------------------------------------------------

. lincom _b[wgt] + _b[frnwgt]

 ( 1)  wgt + frnwgt = 0.0

------------------------------------------------------------------------------
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |  -6.668487   9.821781    -0.68   0.500    -26.27831    12.94133
------------------------------------------------------------------------------

. lincom _b[wgtsq] + _b[frnwgtsq]

 ( 1)  wgtsq + frnwgtsq = 0.0

------------------------------------------------------------------------------
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   .0023235   .0019145     1.21   0.229     -.001499    .0061459
------------------------------------------------------------------------------

. lincom _b[mpg] + _b[frnmpg]

 ( 1)  mpg + frnmpg = 0.0

------------------------------------------------------------------------------
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |  -22.07724   81.23592    -0.27   0.787      -184.27    140.1155
------------------------------------------------------------------------------

. * Equation (4B)
. regress price wgt wgtsq mpg dom domwgt domwgtsq dommpg

------------------------------------------------------------------------------
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         wgt |  -6.668487   9.821781    -0.68   0.500    -26.27831    12.94133
       wgtsq |   .0023235   .0019145     1.21   0.229     -.001499    .0061459
         mpg |  -22.07724   81.23592    -0.27   0.787      -184.27    140.1155
         dom |   2474.355   14701.79     0.17   0.867    -26878.71    31827.42
```

output omitted

Computing domestic-car coefficient estimates from Equation (4B)

```
. * Following Equation (4B)

. lincom _b[_cons] + _b[dom]

 ( 1)  dom + _cons = 0.0

------------------------------------------------------------------------------
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   11972.05   7197.498     1.66   0.101    -2398.214    26342.32
------------------------------------------------------------------------------

. lincom _b[wgt] + _b[domwgt]

 ( 1)  wgt + domwgt = 0.0

------------------------------------------------------------------------------
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |  -8.813044   3.023761    -2.91   0.005    -14.85018    -2.77591
------------------------------------------------------------------------------

. lincom _b[wgtsq] + _b[domwgtsq]

 ( 1)  wgtsq + domwgtsq = 0.0

------------------------------------------------------------------------------
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   .0018982   .0004203     4.52   0.000      .001059    .0027374
------------------------------------------------------------------------------

. lincom _b[mpg] + _b[dommpg]

 ( 1)  mpg + dommpg = 0.0

------------------------------------------------------------------------------
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   78.07451   115.5373     0.68   0.502    -152.6031    308.7521
------------------------------------------------------------------------------

. * Equation (4A)
. regress price wgt wgtsq mpg frn frnwgt frnwgtsq frnmpg

------------------------------------------------------------------------------
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         wgt |  -8.813044   3.023761    -2.91   0.005    -14.85018    -2.77591
       wgtsq |   .0018982   .0004203     4.52   0.000      .001059    .0027374
         mpg |   78.07451   115.5373     0.68   0.502    -152.6031    308.7521
         frn |  -2474.355   14701.79    -0.17   0.867    -31827.42    26878.71
```

output omitted