Printout

Tuesday, February 20, 2018 7:02 AM

6. Extensions of The Two-Variable Linear Regression Mode

So far, $Y_i = \beta_1 + \beta_2 X_i + u_i$

6.1 Functional Form of regression Models

We will consider the following models:

[1] The log-linear model

[2] Semilog models

[3] Reciprocal models

[4] The logarithmic reciprocal model

The Log-linear Model

Consider $Y_i = \beta_1 X_i^{\beta_2} e^{u_i}$ Exponential regression model

where $e = 2.718$

$$\boxed{\ln Y_i = \ln \beta_1 + \beta_2 \ln X_i + u_i}$$

$\ln$ = natural log (i.e., log to the base e where $e = 2.718$.

Let's define $\alpha = \ln \beta_1$. So,

$\ln Y_i = \alpha + \beta_2 \ln X_i + u_i$ → Double log model
→ log-log model
log-linear model

$$\frac{d \ln Y_i}{dx} = \beta_2 \frac{d(\ln x_i)}{dx}$$

$$\frac{1}{Y}\frac{dy}{dx} = \beta_2 \frac{1}{x}\cancel{\frac{dx}{dx}}$$

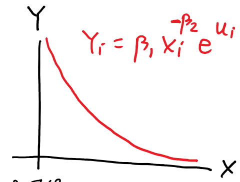
$$\beta_2 = \frac{x}{Y}\frac{dy}{dx} = \frac{\frac{dy}{Y}\times 100}{\frac{dx}{x}\times 100} = \frac{\%\Delta Y}{\%\Delta X} = \text{Elasticity!}$$

NOTE #1
- ΔX = absolute change in X (i.e., $= X_2 - X_1$)
  ΔY = " ——— " Y (i.e., $= Y_2 - Y_1$)
- $\frac{\Delta X}{X}$ = relative change in X (i.e., $= \frac{X_2 - X_1}{X_1}$ → $= \frac{X_2}{X_1} - 1$
- $\frac{\Delta Y}{Y}$ " ——— (i.e., $= \frac{Y_2 - Y_1}{Y_1}$ → $= \frac{Y_2}{Y_1} - 1$
- $\frac{\Delta X}{X} \times 100$ = relative change $\times 100$ = $\boxed{\text{percentage change in } X}$ → $\boxed{\frac{X_2 - X_1}{X_1} \times 100}$ OR $\boxed{\left(\frac{X_2}{X_1} - 1\right) \times 100}$
- $\frac{\Delta Y}{Y} \times 100$ = " ——— " = $\boxed{\text{" ——— " } Y}$ → $\boxed{\frac{Y_2 - Y_1}{Y_2} \times 100}$ OR $\boxed{\left(\frac{Y_2}{Y_1} - 1\right) \times 100}$

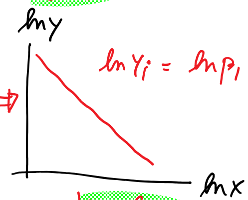
NOTE #2

$$\boxed{\text{SLOPE} = \frac{dY}{dX} = \beta_2 \cdot \frac{Y}{X}}$$

Example



⇒



Y = Quantity demanded for X

X = Price of X

Page 3

Tuesday, February 20, 2018 7:02 AM

The Semilog Models

log-lin model : $\ln Y_i = \beta_1 + \beta_2 X_i + u_i$

lin-log model $Y_i = \beta_1 + \beta_2 \ln X_i + u_i$

① log-linear model (log-lin model)

$$\ln Y_i = \beta_1 + \beta_2 X_i + u_i$$

$$\frac{d \ln Y}{dX} = \beta_2 \cancel{\frac{dX}{dX}}$$

$$\frac{1}{Y}\frac{dY}{dX} \qquad \beta_2$$

$$\beta_2 = \frac{\frac{dY}{Y}}{dX} = \frac{\frac{\Delta Y}{Y}}{\Delta X}$$

$\frac{\Delta Y}{Y}$ → Relative change in Y

ΔX ⟶ Absolute change in X

called " Semi-elasticity of Y w.r.t. X

Q: How to utilize β_2 ?

A: suppose X = Years in school (Schooling)

Y = Wage or salary

$$\ln Y_i = \beta_1 + \beta_2 X_i + u_i \quad \text{(wage-education model)}$$

with our recipe : $\boxed{\beta_2 = \frac{\frac{dY}{Y}}{dX}}$ → $\frac{dY}{Y} = \beta_2 \cdot dX$

$$\frac{dY}{Y} \times 100 = 100 \times \beta_2 \cdot dX$$

$\%\Delta Y$ $\quad \Delta X$

$$* \boxed{\frac{\frac{dY}{Y} \times 100}{dX} = 100 \times \beta_2} *$$

If we multiple β_2 by 100, we obtain $\frac{\%\Delta Y}{\Delta X}$ and we will be able to respond to this type of question : If X changes by 1 unit, Y will change by HOW MANY PERCENT ?

Suppose $\ln Y_i = \beta_1 + \beta_2 X_i + u_i$

salary $\qquad$ years in school

suppose $\hat{\beta}_2 = +0.094$

Then $\%\Delta Y = 100 \cdot \beta_2 \cdot \Delta X = 100 \cdot 0.094 \times 1$

Then $\% \Delta Y = 100 \cdot \beta_2 \cdot \Delta X = 100 \cdot 0{,}094 \times 1$

$= 9.4$

Interpretation: Ceteris paribus, if X increases by 1 unit, Y would increase by 9.4 %

② linear-log model (or lin-log model)

$$Y_i = \beta_1 + \beta_2 \ln X_i + u_i$$

$$\frac{dY}{dX} = \beta_2 \frac{d \ln X_i}{dX}$$

$$\frac{dY}{dX} = \beta_2 \cdot \frac{1}{X} \cancel{\frac{dX}{dX}}$$

$$\beta_2 = X \cdot \frac{dY}{dX}$$

$$\boxed{\beta_2 = \frac{dY}{\frac{dX}{X}}} \rightarrow = \frac{\Delta Y}{\frac{\Delta X}{X}}$$

ΔY → absolute change in Y (i.e., $Y_2 - Y_1$)

$\frac{\Delta X}{X}$ → relative change in X (i.e., $\frac{X_2 - X_1}{X_1}$)

How to utilize this β_2?

$$\beta_2 = \frac{dY}{\frac{dX}{X}}$$

$$dY = \beta_2 \cdot \frac{dX}{X}$$

$$100 \cdot dY = \beta_2 \cdot \frac{dX}{X} \cdot 100$$

$$\boxed{\frac{dY}{\frac{dX}{X} \cdot 100}} = \frac{\beta_2}{100} = 0.01 \cdot \beta_2$$

(type of question: If X changes by 1 %, Y will change by __?__ unit, ceteris paribus

Example $n = 52$ families (in india)

$$\text{FOOD EXP}_i = -1283.912 + 257.27 \ln(\text{TOTAL EXP.}) \quad (257.27 = \hat{\beta}_2)$$

Example   n = 92 families (in India)

$$\text{FOOD EXP}_i = -1283.912 + \overbrace{257.27}^{\beta_2} \ln(\text{TOTAL EXP}_i)$$

(unit: Rupees)

Q: If total expenditure rises by 1%, food expenditure will go up by $\underline{0.01 \times 257.27 = 2.57 \text{ RUPEES}}$.

With lin-log model, what if you want to get SLOPE AND ELASTICITY ?

From $\Delta Y = \beta_2 \cdot \dfrac{\Delta X}{X}$,

SLOPE $\Rightarrow \dfrac{\Delta Y}{\Delta X} = \underline{\dfrac{\beta_2}{X}}$

ELASTICITY $\Rightarrow \dfrac{\%\Delta Y}{\%\Delta X} = \underline{\dfrac{\beta_2}{Y}}$

verify this @ home

Page 4

Tuesday, February 20, 2018 7:02 AM

The Reciprocal Models

Consider $Y_i = \beta_1 + \beta_2 \left[\frac{1}{X_i}\right] + u_i$

- If X_i increases indefinitely, the term $\beta_2 \cdot \frac{1}{X_i}$ will apprach to zero and Y_i will converge to β_1.

- $\frac{dy}{dx} = \beta_2 \frac{d(X_i^{-1})}{dx}$

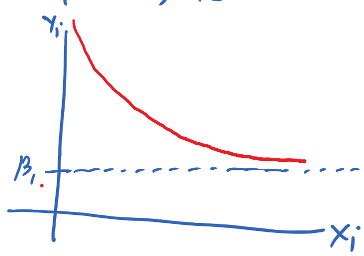
$$\frac{dy}{dx} = -\beta_2 X_i^{-2} = -\beta_2 \cdot \frac{1}{X_i^2}$$

Slope of the curve

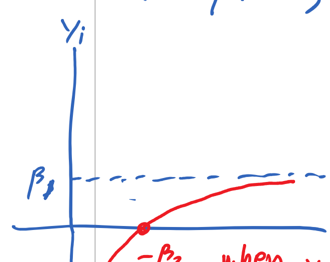
If $\beta_2 > 0$, $\frac{dy}{dx}$ is negative

If $\beta_2 < 0$, $\frac{dy}{dx}$ is positive

Case 1 · $\beta_1 > 0$, $\beta_2 > 0$

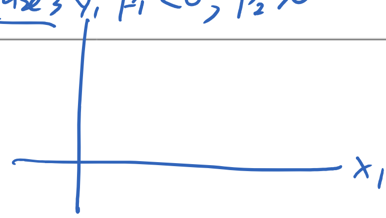


Case 2 $\beta_1 > 0$, $\beta_2 < 0$



when $y = 0$, $X_i = -\frac{\beta_2}{\beta_1}$

Case 3 $\beta_1 < 0$, $\beta_2 > 0$



Case 4 $\beta_1 < 0$, $\beta_2 < 0$

The Logarithmic Reciprocal Models

Page 6

Tuesday, February 20, 2018 7:02 AM

6.2 Regression Through the Origin

LATER

In this section, we consider the case that the two-variable PRF assumes the following form:

$$Y_i = \beta_2 X_i + u_i$$

This model is called **the regression through the origin** where the intercept term $\hat{\beta}_1$ is absent from the model.

Example

Page 7

Tuesday, February 20, 2018 7:02 AM

Since it is the linear regression model, we can apply the Ordinary Least Square (OLS) to estimate the formula for $\hat{\beta}_2$

Let us first write the sample regression function (SRF) as:

$$Y_i = \hat{\beta}_2 X_i + \hat{u}_i$$

We would like to minimize

$$\sum \hat{u}_i{}^2 = \sum (Y_i - \hat{\beta}_2 X_i)^2$$

therefore,

$$\boxed{\hat{\beta}_2 = \frac{\sum X_i Y_i}{\sum X_i^2}}$$

Now we can find out the variance of $\hat{\beta}_2$

$$\mathrm{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum X_i^2}$$

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-1}$$

It should be noted that we get the condition $\sum \hat{u}_i X_i = 0$ from the normal equation. However, with the regression through the origin model, we cannot get the condition $\sum \hat{u}_i = 0$.

For the zero-intercept model, r^2 can be negative, whereas for the conventional model it cannot be negative.

Since the conventional r^2 is not appropriate for the regressions that do not contain the intercept, we therefore compute what is known as the **raw** r^2 instead:

$$\text{raw}\, r^2 = \frac{(\sum X_i Y_i)^2}{\sum X_i^2 \sum Y_i^2}$$

This raw r^2 has its value between 0 and 1, but we cannot directly compare its value to the conventional r^2 value. For this reason, some researchers do not report the r^2 value for zero intercept regression models.

This raw r^2 has its value between 0 and 1, but we cannot directly compare its value to the conventional r^2 value. For this reason, some researchers do not report the r^2 value for zero intercept regression models.

rescaled Y_i

$Y_i^* = \frac{Y_i}{1000}$

or $Y_i^* = w_2 \cdot Y_i$

where $w_2 = \frac{1}{1000}$

6.2.1 Scaling and Units of Measurements

Consider our old example given in table 18 which refer to weekly family expenditure (Y) and Income (x), in baht.

Table 5. A Random Sample From the Population

X	Y
500	390
600	425
700	560
800	575
900	630
1000	679

X: baht
Y: baht

By using the OLS estimation, we get the following results:

$$\hat{Y}_i = \underbrace{98.524}_{\hat{\beta}_1} + \underbrace{0.593}_{\hat{\beta}_2} X_i$$ (All measured in baht)

- $var(\hat{\beta}_1) = 2706.3712 \rightarrow se(\hat{\beta}_1) = 52.023$
- $var(\hat{\beta}_2) = 0.0046 \rightarrow se(\hat{\beta}_2) = 0.0678$
- residual variance $\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-2} = 800.476$

Now, we are interested in changing the units of our data. For example, we would prefer to express our sample data in the unit of 1000 baht. By using the new unit of X and Y, we can report our data in 1000 baht as in the following table.

Table 5. A Random Sample From the Population

X	Y
500	390
600	425
700	560
800	575
900	630
1000	679

0.39
0.425
0.56
0.575
0.630
0.679

6.2 Regression Through the Origin

Table 6.2: Weekly family Expe

X	Y
0.5	0.360
0.6	0.390
0.7	0.440
0.8	0.575
0.9	0.670
1	0.730

$X_i^* = \frac{X_i}{1000}$

$X_i^* = w_1 X_i$

where $w_1 = \frac{1}{1000}$

With the new unit, we would like to answer these two questions:
1. Do the units in which the regressand (Y) and regressor/s (X) are measured make any difference in the regression results?
2. If so, what is the sensible course to follow in choosing units of measurement for regression analysis?

To answer these questions, let:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i$$

where Y is the weekly family expenditure and X is the income, in baht.

Now, let w_1 and w_2 are constants, called the **Scale factors**. For example, in our data, if we need to use the unit of 1000 baht instead, we can directly multiple the original data in table 18 with the scale factors equal to 0.001. In other words, $w_1 = w_2 = \frac{1}{1000} = 0.001$.

Define

$$\boxed{Y_i^* = w_1 Y_i}$$

$$\boxed{X_i^* = w_2 X_i}$$

$$\boxed{w_1 = w_2 = \frac{1}{1000}}$$

Now consider the regression using Y_i^* and X_i^* variables:

$$Y_i^* = \hat{\beta}_1^* + \hat{\beta}_2^* X_i^* + \hat{u}_i^* \rightarrow \text{RESCALED MODEL}$$

$\hat{u}_i^* = ?$

LET'S COMPARE ORIGINAL MODEL WITH RESCALED MODEL

ORIGINAL MODEL

$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}$

$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2}$

$var(\hat{\beta}_1) = \frac{\sum X_i^2}{n \cdot \sum x_i^2} \cdot \sigma^2$

$var(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2}$

RESCALED MODEL

$\hat{\beta}_1^* = \bar{Y}^* - \hat{\beta}_2^* \bar{X}^*$

$\hat{\beta}_2^* = \frac{\sum x_i^* y_i^*}{\sum x_i^{*2}}$

$var(\hat{\beta}_1^*) = \frac{\sum X_i^{*2}}{n \cdot \sum x_i^{*2}} \cdot \sigma^{*2}$

$var(\hat{\beta}_2^*) = \frac{\sigma^{*2}}{}$

$$n \cdot \sum x_i^{*2}$$

$$\operatorname{var}(\hat\beta_2) = \frac{\sigma^2}{\sum x_i^2} \qquad \operatorname{var}(\hat\beta_2^*) = \frac{\sigma^{*2}}{\sum x_i^{*2}}$$

$$\hat\sigma^2 = \frac{\sum \hat u_i^2}{n-2} \qquad \hat\sigma^{*2} = \frac{\sum \hat u_i^{*2}}{n-2}$$

where

$$Y_i^* = w_1 Y_i \quad \text{or} \quad y_i^* = w_1 \cdot y_i$$

$$X_i^* = w_2 X_i \quad \text{or} \quad x_i^* = w_2 \cdot x_i$$

$$\hat u_i^* = w_1 \cdot \hat u_i$$

$$\bar Y^* = w_1 \bar Y$$

$$\bar X = w_2 \bar X$$

Page 12

Tuesday, February 20, 2018 7:02 AM

Our target is to find out the relationship between the following pairs:

1. $\hat{\beta}_1$ and $\hat{\beta}_1^*$

2. $\hat{\beta}_2$ and $\hat{\beta}_2^*$

3.var($\hat{\beta}_1$) and var($\hat{\beta}_1^*$)

4.var($\hat{\beta}_2$) and var($\hat{\beta}_2^*$)

5.$\hat{\sigma}^2$ and $\hat{\sigma}^{*2}$

6. r_{xy}^2 and $r_{x^*y^*}^2$

1. $\hat{\beta}_1$ and $\hat{\beta}_1^*$

$$\hat{\beta}_1^* = w_1 \cdot \hat{\beta}_1$$

2. $\hat{\beta}_2$ and $\hat{\beta}_2^*$

$$\hat{\beta}_2^* = \left(\frac{w_1}{w_2}\right) \cdot \hat{\beta}_2 \qquad \frac{1}{\frac{1}{20}} \to 20$$

in our case

So, ... if $w_1 = w_2$, $\hat{\beta}_2^* = \hat{\beta}_2$

X	Y
BAHT	BAHT
1000 BAHT	1000 BAHT

$w_1 = \frac{1}{1000}$, $w_2 = \frac{1}{1000}$

3. $\text{var}(\hat{\beta}_1)$ and $\text{var}(\hat{\beta}_1^*)$

$$\text{var}(\hat{\beta}_1^*) = w_1^2 \cdot \text{var}(\hat{\beta}_1)$$

4. $\text{var}(\hat{\beta}_2)$ and $\text{var}(\hat{\beta}_2^*)$

$$\text{var}(\hat{\beta}_2^*) = \left(\frac{w_1}{w_2}\right)^2 \cdot \text{var}(\hat{\beta}_2)$$

Again, if $w_1 = w_2$, $\text{var}(\hat{\beta}_2^*) = \text{var}(\hat{\beta}_2)$

$se(\hat{\beta}_2^*) = se(\hat{\beta}_2)$

5. $\hat{\sigma}^2$ and $\hat{\sigma}^{*2}$

$$\boxed{\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-2}} \quad \text{Vs.} \quad \boxed{\hat{\sigma}^{*2} = \frac{\sum \hat{u}_i^{*2}}{n-2}}$$

Where $\hat{u}_i^* = w_1 \cdot \hat{u}_i$

6. r^2_{xy} and $r^2_{x^*y^*}$

$r^2_{xy} = r^2_{x^*y^*}$

where $x_i^* = w_2 \cdot x_i$, $y_i^* = w_1 \cdot y_i$

NOTE: $x = X_i - \bar{X}$, $y_i = Y_i - \bar{Y}$

R-squared does not change.

Example:

BWGHT = Child birth weight, in ounces

CIG = Number of cigarettes smoked by the mother while pregnant, per day

FAMINC = Annual family income, in thousand of dollars

$$\boxed{\widehat{BWGHT} = \hat{\beta}_0 + \hat{\beta}_1 CIG + \hat{\beta}_2 FAMINC}$$

$= \frac{BWGHT}{16}$

Effect of Data Scaling

Dependent variable / Independent variables	(1) BWGHT	(2) BWGHTLBS	(3) BWGHT
CIG	-.4634 (0.0916) S.D	-0.0289 (.0057) S.D	—
PACKS	—	—	-9.268
FAMINC	.0927	.0058	.0927
INTERCEPT	116.974	7.3109	116.974
OBSERVATION	1388	1388	1388
R^2	0.298	0.298	0.298

$\hat{t} = \frac{COEFF}{SD}$

$\hat{t} = 5.0589$ very statistically significant

- If a mother smokes 5 cigarettes per day, on average, weight is predicted to be about $0.4634 \times 5 = 2.317$ ounces LESS!
- If we measure child birth weight in pounds instead:

$$BWGHTLBS = \frac{BWGHT}{16}$$

measured in pounds.

Result: $\frac{\widehat{BWGHT}}{16} = \hat{\beta}_0/16 + (\hat{\beta}_1/16)\, CIG + (\hat{\beta}_2/16)\, FAMINC$

TSS = ESS + RSS

BWGHTLBS

NO CHANGE

Dependent Variable: BWGHT → OUNCES
Method: Least Squares
Date: 03/22/18 Time: 09:59
Sample: 1 1388
Included observations: 1388

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	116.9741	1.048984	111.5118	0.0000
CIGS	-0.463408	0.091577	-5.060315	0.0000
FAMINC	0.092765	0.029188	3.178195	0.0015

R-squared	0.029805	Mean dependent var	118.6996
Adjusted R-squared	0.028404	S.D. dependent var	20.35396
S.E. of regression	20.06282	Akaike info criterion	8.837772
Sum squared resid	557485.5	Schwarz criterion	8.849089
Log likelihood	-6130.414	Hannan-Quinn criter.	8.842005
F-statistic	21.27392	Durbin-Watson stat	1.921690
Prob(F-statistic)	0.000000		

ESS

RSS

$\Sigma \hat{u}_i^2$

$\times \frac{1}{16}$ $\times \frac{1}{16}$ $\times \frac{1}{16}$ $\times \frac{1}{256}$

Dependent Variable: BWGHTLBS → POUNDS
Method: Least Squares
Date: 03/22/18 Time: 10:10
Sample: 1 1388
Included observations: 1388

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.310883	0.065562	111.5118	0.0000
CIGS	-0.028963	0.005724	-5.060315	0.0000
FAMINC	0.005798	0.001824	3.178195	0.0015

R-squared	0.029805	Mean dependent var	7.418723
Adjusted R-squared	0.028404	S.D. dependent var	1.272123
S.E. of regression	1.253926	Akaike info criterion	3.292595
Sum squared resid	2177.678	Schwarz criterion	3.303911
Log likelihood	-2282.061	Hannan-Quinn criter.	3.296827
F-statistic	21.27392	Durbin-Watson stat	1.921690
Prob(F-statistic)	0.000000		

$\hat{t} = \frac{\hat{coff}/16}{SE/16}$

$\left(\frac{\hat{u}_i}{16}\right)^2 = \left(\frac{\hat{u}_i^2}{256}\right)$

$\widehat{BWGHT} = \hat{\beta}_0 + \hat{\beta}_1 CIGS + \hat{\beta}_2 FAMINC$

$\hat{\beta}_1 = -0.4634$

$\frac{\widehat{BWGHT}}{16} = \frac{\hat{\beta}_0}{16} + \frac{\hat{\beta}_1}{16} CIGS + \frac{\hat{\beta}_2}{16} FAMINC$

$\frac{\hat{\beta}_1}{16} = \frac{-0.4634}{16} = -0.02896.$

Ceteris paribus, if an extra cig being smoked, child birth weight will

Ceteris paribus, if an extra cig being

child birth weight, decrease by 0.4634 ounces.

If 5 CIGS, weight ↓ by

$= 5(-0.4634)$

$= 2.317$ ounces

consumed, child birth weight will decrease by 0.02896 pounds

If 5 CIGS, weight ↓ by

$= 5(-0.02896)$

$= 0.1445$ pounds

$\times 16$

Changing unit of measurement does not affect the result. It just change the way to report it

In other word, the same answer regardless of how the dependent variable is measured!

$\boxed{\text{Model (3)}}$

$$\widehat{BWGHT} = \hat{\beta}_0 + 20 \cdot \hat{\beta}_1 (CIG/20) + \hat{\beta}_2 \, FAMINC$$

$$= \quad + 20 \cdot \hat{\beta}_1 (PACKS) + \hat{\beta}_2 \, FAMINC$$

$\hat{\beta}_1$

$= 20 \cdot (-0.4634)$

Dependent Variable: BWGHT
Method: Least Squares
Date: 03/22/18 Time: 10:46
Sample: 1 1388
Included observations: 1388

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	116.9741	1.048984	111.5118	0.0000
PACKS	-9.268151	1.831536	-5.060315	0.0000
FAMINC	0.092765	0.029188	3.178195	0.0015

R-squared	0.029805	Mean dependent var	118.6996
Adjusted R-squared	0.028404	S.D. dependent var	20.35396
S.E. of regression	20.06282	Akaike info criterion	8.837772
Sum squared resid	557485.5	Schwarz criterion	8.849089
Log likelihood	-6130.414	Hannan-Quinn criter.	8.842005
F-statistic	21.27392	Durbin-Watson stat	1.921690
Prob(F-statistic)	0.000000		

+5 CIGS → weight reduces by

$$\frac{-9.268151}{4} = -2.317 \text{ ounces}$$