Econ 103: Introduction to Econometrics
Week 3

Francisco Díaz-Valdés*

Fall 2025

1 Functional form

The starting point in all econometric analyses is economic theory. What does economics say about the relation between food expenditure and income, holding all else constant? We expect a positive relationship between these variables because food is a normal good. But nothing says the relationship must be a straight line. We do not expect that as household income rises, food expenditures will continue to increase indefinitely at the same constant rate. Instead, as income increases, we expect food expenditures to rise, but we expect such expenditures to increase at a decreasing rate. In the economic context of the food expenditure model, the marginal propensity to spend on food is greater at lower incomes, and as income increases, the marginal propensity to spend on food declines.

In light of the previous paragraphs, we might need to transform the variables of the model while keeping the model **linear in the coefficients**. Below is a list of non-linear models. Of course, you can transform Y_i or X_i and run OLS, so which model you use will depend on the context.

$$Y_i = \beta_1 + \beta_2 X_i^2 + \varepsilon_i \qquad \text{(Quadratic Model)}$$
$$Y_i = \beta_1 + \beta_2 X_i^3 + \varepsilon_i \qquad \text{(Cubic Model)}$$
$$Y_i = \beta_1 + \beta_2 \log(X_i) + \varepsilon_i \qquad \text{(Linear-log Model)}$$
$$\log(Y_i) = \beta_1 + \beta_2 X_i + \varepsilon_i \qquad \text{(Log-linear Model)}$$
$$\log(Y_i) = \beta_1 + \beta_2 \log(X_i) + \varepsilon_i \qquad \text{(Log-log Model)}$$

1.1 Interpreting the coefficients

Recall that the slope for any model is given by: $\partial \mathbb{E}[y \mid x]/\partial x$ In the simple model, $y = \beta_1 + \beta_2 x + \varepsilon$, the interpretation is straightforward: $\beta_2 = \partial \mathbb{E}[y \mid x]/\partial x$. However, don't rush to interpret the effect of a change in x as just β_2 in other models. The interpretation is somewhat different for the log-log, log-linear, and linear-log models. In the TA session, we will discuss how to interpret the coefficients for these models, including the respective intuition.

	Model	Slope
Linear	$Y = \beta_1 + \beta \cdot X + \varepsilon$	β_2
Quadratic	$Y = \beta_1 + \beta_2 \cdot X^2 + \varepsilon$	$2\beta_2 x$
Cubic	$Y = \beta_1 + \beta_2 \cdot X^3 + \varepsilon$	$3\beta_2^2 x$

Table 1: Examples of models and slopes

*Many thanks to all previous TAs for providing the notes. All mistakes are my own. Please get in touch with me at fdiazvaldes@g.ucla.edu if you spot any typos or mistakes.

1.2 Testing for non-linearities (totally optional)

The Ramsey Regression Equation Specification Error Test is a general specification test for the linear regression model. More specifically, it tests whether non-linear combinations of the explanatory variables help to explain the response variable.
The null model is

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$$

which is estimated by OLS, yielding predicted values $\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i$. Now let

$$z_i = \begin{pmatrix} \widehat{y}_i^2 \\ \vdots \\ \widehat{y}_i^k \end{pmatrix}$$

be a $(k-1)$-vector of powers of $\hat{y}_i$. Run the auxiliary regression

$$y_i = \beta_1 + \beta_0 x_i + \gamma_1 \hat{y}_i^2 + \ldots + \gamma_{k-1} \hat{y}_i^k + \varepsilon_i$$

By OLS, and form the Wald statistic W_T for $\mathrm{H}_0 : \gamma_1 = \ldots = \gamma_{k-1} = 0$. It can be shown that under the null hypothesis, $W_T \sim \chi^2_{k-1}$. Thus, the null hypothesis is rejected at the $\alpha\%$ level if W_T exceeds the upper $\alpha\%$ tail critical value of the χ^2_{k-1} distribution. If the null hypothesis that all γ coefficients are zero is rejected, then the model suffers from misspecification. To implement the test, k must be selected in advance. Typically, small values such as $k = 2, 3$ or 4 seem to work best.

2 Confidence intervals

Let's assume that assumptions SR1-SR6 hold for the simple linear regression model. In this case, we know that given $\mathbf{x}$, the OLS estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ have normal distributions. For example, the normal distribution of $\hat{\beta}_2$ is

$$\hat{\beta}_2 \mid \mathbf{x} \sim \mathcal{N}\left(\beta_2, \frac{\sigma^2}{\sum\left(x_i-\bar{x}\right)^2}\right)$$

From there, we can get the standardized normal random variable Z for $\hat{\beta}_2$ as:

$$Z=\frac{\hat{\beta}_2-\beta_2}{\sqrt{\sigma^2 / \sum\left(x_i-\bar{x}\right)^2}} \sim \mathcal{N}(0,1)$$

From your introduction to statistics course, you know that:

$$\mathbb{P}(-1.96 \leq Z \leq 1.96)=0.95$$

Manipulating this last expression, we can get:

$$\begin{aligned}
& \mathbb{P}\left(-1.96 \leq \frac{\hat{\beta}_2-\beta_2}{\sqrt{\sigma^2 / \sum\left(x_i-\bar{x}\right)^2}} \leq 1.96\right)=0.95 \\
\Leftrightarrow \quad & \mathbb{P}\left(\hat{\beta}_2-1.96 \sqrt{\sigma^2 / \sum\left(x_i-\bar{x}\right)^2} \leq \beta_2 \leq \hat{\beta}_2+1.96 \sqrt{\sigma^2 / \sum\left(x_i-\bar{x}\right)^2}\right)=0.95
\end{aligned}$$

This defines an interval that has a probability of 0.95 of containing the parameter β_2. However, we don't know the parameter σ^2, which also has to be estimated. Recall that $\hat{\varepsilon}_i=y_i-(\hat{\beta}_1-\hat{\beta}_2 x_i)$, the estimator of σ^2 is $\hat{\sigma}^2=\sum_i \hat{\varepsilon}^2/(N-2)$. By using $\hat{\sigma}^2$, we are changing the distribution of the statistic from a normal distribution to a t-student distribution with $N-2$ degrees of freedom,

$$t=\frac{\hat{\beta}_2-\beta_2}{\sqrt{\hat{\sigma}^2 / \sum\left(x_i-\bar{x}\right)^2}}=\frac{\hat{\beta}_2-\beta_2}{\sqrt{\widehat{\operatorname{var}}\left(\hat{\beta}_2\right)}}=\frac{\hat{\beta}_2-\beta_2}{\operatorname{se}\left(\hat{\beta}_2\right)} \sim t_{(N-2)}$$

It looks like the standard normal distribution, except it is more spread out, with a larger variance and thicker tails. A single parameter controls the shape of the t-distribution called the degrees of freedom, often abbreviated as df. However, as $N \rightarrow \infty$ the t-student distribution converges to a $\mathcal{N}(0,1)$.

We can find a "critical value" t_c from a t-distribution such that $\mathbb{P}\left(t \geq t_c\right)=P\left(t \leq-t_c\right)=\alpha / 2$, where α is a probability often taken to be $\alpha=0.01$ or $\alpha=0.05$. That is, we are looking for t_c such that:

$$\mathbb{P}(-t_c \leq t \leq t_c)=1-\alpha$$

This last statement relies on the t-student distribution being symmetrical for $df>3$.
Suppose that $\alpha=0.05$ when we are looking for:

$$\mathbb{P}\left[-t_{(0.975, N-2)} \leq t \leq t_{(0.975, N-2)}\right]=0.95$$

Putting all these bits together, we got:

$$\begin{aligned}
& \mathbb{P}\left[-t_c \leq \frac{\hat{\beta}_k-\beta_k}{\operatorname{se}\left(\hat{\beta}_k\right)} \leq t_c\right]=1-\alpha \\
\Leftrightarrow \quad & \mathbb{P}\left[\hat{\beta}_k-t_c \operatorname{se}\left(\hat{\beta}_k\right) \leq \beta_k \leq \hat{\beta}_k+t_c \operatorname{se}\left(\hat{\beta}_k\right)\right]=1-\alpha
\end{aligned}$$

Practice Questions

Question 1

Let the simple regression model $Y = \beta_1 + \beta_2 \cdot X + e$. Consider the inference that tests the null hypothesis $H_0 : \beta_2 = 0$ against $H_1 : \beta_2 \neq 0$ at significance level α. Which of the statements is false?

(a) If the standard error of $\hat{\beta}_2$ decreases, then it is more likely to reject H_0, (everything else constant).

(b) The higher the absolute value of $\hat{\beta}_2$, the more likely it is to reject H_0 (everything else constant).

(c) The higher the significance level α, the more likely it is to reject H_0 (everything else constant).

(d) Hypothesis $H_0 : \beta_2 = 0$ is not rejected whenever the value 0 belongs to its confidence interval (with a confidence level of $1 - \alpha$).

(e) The larger the sample size, the more likely it is to reject H_0 (everything else constant).

(f) The higher the p-value, the more likely you are to reject H_0.

Answer: the correct answer is (f).

Question 2

Consider the following regression model:

$$Y_i = \beta_1 + \beta_2 x_i + \epsilon_i,$$

for $i = 1, \ldots, N$. Let $e_i \sim \mathcal{N}\left(0, \sigma^2\right)$. That is, e_i has a distribution whose mean is 0, and its variance is σ^2. Let

$$s_x^2 = \frac{1}{N} \sum_{i=1}^{N} \left(x_i - \bar{x}\right)^2,$$

where

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

One ran a regression of y on x and obtained the following estimates for β_1 and $\beta_2 : \hat{b}_1 = 4, \hat{b}_2 = .5$. Define $x_i^* = 10 \times x_i$. If one were to run a regression of y on x^* the estimates $\hat{b}_1^*$ and $\hat{b}_2^*$ would be

(a) $\hat{b}_1^* = 4, \hat{b}_2^* = 5$

(b) $\hat{b}_1^* = 4, \hat{b}_2^* = .5$

(c) $\hat{b}_1^* = 4, \hat{b}_2^* = .05$

(d) $\hat{b}_1^* = .4, \hat{b}_2^* = .5$

(e) $\hat{b}_1^* = .4, \hat{b}_2^* = 5$

Answer: the correct answer is (c).