Regression analysis with cross-sectional data: Confidence intervals for predictions

Bernard Lejeune

Supplemental lecture notes II

Advanced Econometrics
HEC-University of Liège
Academic year 2021-2022

- These lecture notes restate, in matrix form and with more details, the content of Section 6.4a of Wooldridge (2016).

1. Predictions

- Let the $1 \times k$ (row) vector $X_0 = (1, x_{02}, ..., x_{0k})$ denotes the (fixed, nonstochastic) value of explanatory variables for which we want to make predictions. Two different predictions (or forecasts) must be distinguished:
  - A prediction of the expected value of y given X_0: $E(y_0|X_0) = X_0\beta$
  - A prediction of the value of y given X_0: $y_0 = X_0\beta + u_0$

1.1. Prediction of the expected value of y given X_0

- An obvious estimator/predictor of $E(y_0|X_0) = X_0\beta$ is obtained by replacing β by its OLS estimator $\hat{\beta}$:

$$\hat{y}_0 = X_0\hat{\beta}$$

- Under the Gauss-Markov assumptions, we have:

$$E(\hat{y}_0|X) = E(X_0\hat{\beta}|X) = X_0E(\hat{\beta}|X) = X_0\beta$$

which means that $\hat{y}_0$ is a unbiased estimator/predictor of $E(y_0|X_0) = X_0\beta$, and:

$$Var(\hat{y}_0|X) = Var(X_0\hat{\beta}|X) = X_0V(\hat{\beta}|X)X_0' = \sigma^2X_0\left(X'X\right)^{-1}X_0'$$

- Because $E(y_0|X_0) = X_0\beta$ is nothing but a (nonstochastic) single linear combination of β, from the results on confidence intervals outlined in the supplemental lecture notes I (hereafter SLN-I), a $(1-\alpha) \times 100\%$ confidence interval for $E(y_0|X_0) = X_0\beta$ is simply given by:

$$\left[\hat{y}_0 - t_{n-k;1-\frac{\alpha}{2}} s.\hat{e}.(\hat{y}_0)\,;\; \hat{y}_0 + t_{n-k;1-\frac{\alpha}{2}} s.\hat{e}.(\hat{y}_0)\right]$$

where $s.\hat{e}.(\hat{y}_0) = s.\hat{e}.(X_0\hat{\beta}) = \sqrt{V\hat{a}r(X_0\hat{\beta})} = \sqrt{X_0\hat{V}(\hat{\beta})X_0'} = \sqrt{\hat{s}^2 X_0(X'X)^{-1}X_0'}$.
This confidence interval is exact in finite sample if normality holds, and is asymptotically valid – i.e., approximately valid for n sufficiently large – without the normality assumption.

- Remark: It is in practice possible to directly obtain $\hat{y}_0$ and $s.\hat{e}.(\hat{y}_0)$ by running a regression where the explanatory variables are appropriately centered. For details, see Wooldridge (2016), Section 6.4a.

1.2. Prediction of the value of y given X_0

- Because u_0 can not be predicted and has zero mean, the best predictor of $y_0 = X_0\beta + u_0$ is still:

$$\hat{y}_0 = X_0\hat{\beta}$$

- Let $\hat{e}_0 = y_0 - \hat{y}_0 = X_0(\beta - \hat{\beta}) + u_0$ denotes the prediction error in using $\hat{y}_0$ to predict y_0. Under the Gauss-Markov assumptions, we have:

$$\begin{aligned} E(\hat{e}_0|X) &= E\left[(X_0(\beta-\hat{\beta}) + u_0)|X\right] \\ &= X_0\beta - X_0E(\hat{\beta}|X) + E(u_0|X) \\ &= X_0\beta - X_0\beta + 0 = 0 \end{aligned}$$

which means that the prediction error has zero mean (although $E(\hat{y}_0|X) \neq y_0$), and:

$$\begin{aligned} Var(\hat{e}_0|X) &= Var\left[(X_0(\beta-\hat{\beta}) + u_0)|X\right] \\ &= Var(X_0(\beta-\hat{\beta})|X) + 2Cov(X_0(\beta-\hat{\beta}), u_0|X) \\ &\quad + Var(u_0|X) \end{aligned}$$

where[1]:

$$\begin{aligned} Var(X_0(\beta-\hat{\beta})|X) &= Var(-X_0\hat{\beta}|X) = X_0V(\hat{\beta}|X)X_0' \\ &= \sigma^2 X_0\left(X'X\right)^{-1}X_0' \end{aligned}$$

[1] Note that from Section 3.1.1 in SLN-I, we have: $\hat{\beta} - \beta = \left(X'X\right)^{-1}X'u$.

$$\begin{aligned}
Cov(X_0(\beta-\hat{\beta}),u_0|X) &= E(-X_0\,(X'X)^{-1}\,X'uu_0|X) \\
&= -X_0\,(X'X)^{-1}\,X'E\,(uu_0|X) \\
&= -X_0\,(X'X)^{-1}\,X'\cdot 0 = 0 \\
Var(u_0|X) &= \sigma^2
\end{aligned}$$

so that :

$$\begin{aligned}
Var(\hat{e}_0|X) &= Var(\hat{y}_0|X)+Var(u_0|X) \\
&= X_0V(\hat{\beta}|X)X_0'+\sigma^2 \\
&= \sigma^2X_0\,(X'X)^{-1}\,X_0'+\sigma^2
\end{aligned}$$

- If in addition normality is assumed, because $\hat{e}_0$ is a linear function of $\hat{\beta}$ and u_0 which are both normal, then $\hat{e}_0$ is also normally distributed[2] :

$$\hat{e}_0|X \sim N\left(0, Var(\hat{e}_0|X)\right)$$

so that, conditional on X, we have :

$$\hat{z}=\frac{\hat{e}_0}{s.e.(\hat{e}_0|X)}\sim N(0,1) \tag{1}$$

where $s.e.(\hat{e}_0|X)=\sqrt{Var(\hat{e}_0|X)}=\sqrt{\sigma^2X_0\,(X'X)^{-1}\,X_0'+\sigma^2}$.

- Following the same reasoning as in SLN-I, under the same assumptions, conditional on X, we have that $\hat{v}=\frac{(n-k)\hat{s}^2}{\sigma^2}\sim\chi^2(n-k)$ and that $\hat{z}$ and $\hat{v}$ are independent, so that from the definition of the Student distribution, still conditional on X, we have :

$$\hat{t}=\frac{\hat{z}}{\sqrt{\frac{\hat{v}}{n-k}}}=\frac{\frac{\hat{e}_0}{\sqrt{\sigma^2X_0(X'X)^{-1}X_0'+\sigma^2}}}{\sqrt{\frac{\hat{s}^2}{\sigma^2}}}=\frac{\hat{e}_0}{\sqrt{\hat{s}^2X_0\,(X'X)^{-1}\,X_0'+\hat{s}^2}}\sim t(n-k)$$

i.e. :

$$\hat{t}=\frac{\hat{e}_0}{s.\hat{e}.(\hat{e}_0)}=\frac{y_0-\hat{y}_0}{s.\hat{e}.(\hat{e}_0)}\sim t(n-k) \tag{2}$$

where $s.\hat{e}.(\hat{e}_0)=\sqrt{V\hat{a}r(\hat{e}_0)}=\sqrt{\hat{s}^2X_0\,(X'X)^{-1}\,X_0'+\hat{s}^2}=\sqrt{X_0\hat{V}(\hat{\beta})X_0'+\hat{s}^2}$.

In words, if the unknown variance σ^2 appearing in the standard error $s.e.(\hat{e}_0|X)$ of statistic (1) is replaced by its unbiased estimator $\hat{s}^2$, so that the standard error $s.e.(\hat{e}_0|X)$ is replaced by its estimator $s.\hat{e}.(\hat{e}_0)$, then the distribution of (1) switches from normal to Student.

- The distributional result (2) holds conditional on X. But as the conditional distribution of $\hat{t}=\frac{\hat{e}_0}{s.\hat{e}.(\hat{e}_0)}$ actually does not depend on X, it also holds uncondi-

[2] If normality does not hold, $\hat{\beta}$ is still approximately normally distributed, but is is not the case of u_0, so that the distribution of $\hat{e}_0$ is no longer normally distributed, even approximately.

tionally, and we can write:

$$\mathbb{P}\left(-t_{n-k;1-\frac{\alpha}{2}} \leq \frac{y_0 - \hat{y}_0}{s.\hat{e}.(\hat{e}_0)} \leq t_{n-k;1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

so that we have:

$$\mathbb{P}\left(\hat{y}_0 - t_{n-k;1-\frac{\alpha}{2}} s.\hat{e}.(\hat{e}_0) \leq y_0 \leq \hat{y}_0 + t_{n-k;1-\frac{\alpha}{2}} s.\hat{e}.(\hat{\beta}_j)\right) = 1 - \alpha$$

and a $(1-\alpha) \times 100\%$ confidence interval for $y_0 = X_0\beta + u_0$ is given by:

$$\left[\hat{y}_0 - t_{n-k;1-\frac{\alpha}{2}} s.\hat{e}.(\hat{e}_0)\,;\, \hat{y}_0 + t_{n-k;1-\frac{\alpha}{2}} s.\hat{e}.(\hat{e}_0)\right]$$

This confidence interval is exact in finite sample. It requires normality: it does not hold – even approximately – without the normality assumption.

- Remark: For easy computation, note that $s.\hat{e}.(\hat{e}_0) = \sqrt{s.\hat{e}.(\hat{y}_0)^2 + \hat{s}^2}$ and remind that, as outlined above, $\hat{y}_0$ and $s.\hat{e}.(\hat{y}_0)$ may directly be obtained by running a regression where the explanatory variables are appropriately centered. For details, see Wooldridge (2016), Section 6.4a.

Reference

Wooldridge J.M. (2016), *Introductory Econometrics: A Modern Approach*, 6th Edition, Cengage Learning.