Bielefeld University · Econometrics · Summer 2024

Statistical and Econometric Models

Lectures: Lennart Oelschläger · Tutorials: Sebastian Büscher ✉

Problem Set – Week 3 (for tutorial on May 06, 2024)
Solutions

Properties of "Hat Matrix" and "Residual Maker"

Problem 3.1:

In the linear regression framework $y = X\beta + \varepsilon$,

- the "Hat Matrix" H yields the fitted values $\hat{y} = X\hat{\beta}$
- and the "Residual Maker" matrix M yields the residuals $e = y - \hat{y}$

when right-multiplied with y, respectively.

(a) Show that $H = X(X'X)^{-1}X'$ and $M = I - H$.

Solution:

$$
\begin{aligned}
& Hy = \hat{y} = X\hat{\beta} = X(X'X)^{-1}X'y \\
\implies & H = X(X'X)^{-1}X'
\end{aligned}
$$

$$
\begin{aligned}
& My = e = y - \hat{y} = y - X\hat{\beta} = y - X(X'X)^{-1}X'y = (I - H)y \\
\implies & M = I - H
\end{aligned}
$$

(b) Show that H and M are square and symmetric.

Solution: both matrices are $N \times N$, which follows from the dimensions of the matrices $(X \in \mathbb{R}^{N\times(K+1)})$ and their definitions

$$
\begin{aligned}
H' &= (X(X'X)^{-1}X')' && \mid \quad (AB)' = B'A' \\
&= ((X'X)^{-1}X')'X' \\
&= X((X'X)^{-1})'X' && \mid \quad (A^{-1})' = (A')^{-1} \\
&= X((X'X)')^{-1}X' \\
&= X(X'X)^{-1}X' \\
&= H
\end{aligned}
$$

$$
M' = (I - H)' = I - H' = I - H = M
$$

(c) Show that H and M are idempotent.

Solution:

$$\begin{aligned} H^2 &= \left(X(X'X)^{-1}X'\right)\left(X(X'X)^{-1}X'\right) \\ &= X\left((X'X)^{-1}(X'X)\right)(X'X)^{-1}X' \\ &= XI(X'X)^{-1}X' \\ &= X(X'X)^{-1}X' \\ &= H \end{aligned}$$

$$\begin{aligned} M^2 &= (I-H)(I-H) \\ &= I - IH - HI + H^2 \\ &= I - 2H + H^2 && \mid \quad H^2 = H \\ &= I - 2H + H \\ &= I - H \\ &= M \end{aligned}$$

(d) Show that $MX = 0 \in \mathbb{R}^{N\times(K+1)}$.

Solution:

$$\begin{aligned} MX &= (I-H)X \\ &= X - X(X'X)^{-1}X'X \\ &= X - XI \\ &= 0_{N\times(K+1)} \end{aligned}$$

(e) Show that the residuals e are orthogonal to all covariates x_j, i.e. $x_j'e = 0$, where x_j is the j-th column of X.

Solution:

$$\begin{aligned} x_j'e &= x_j'My \\ &= x_j'M'y \\ &= (Mx_j)'y \\ &= 0_{1\times N}y \\ &= 0 \end{aligned}$$

(f) Show that a symmetric and idempotent matrix $A \in \mathbb{N} \times \mathbb{N}$ is positive semi-definite.

Solution: Two ways to show this:

(1) Show it directly. Let $x \in \mathbb{R}^{N \times 1}$, then

$$\begin{aligned} x'Ax &= x'(Ax) \\ &= x'A(Ax) \\ &= x'A'(Ax) \\ &= (Ax)'(Ax) \end{aligned}$$

(2) Showing $\lambda_i \geq 0$ for all eigenvalues λ_i of an idempotent matrix A. Let x be an eigenvector of A with eigenvalue λ, then

$$Ax = \lambda x,$$

but also

$$\begin{aligned} Ax &= (AA)x \\ &= A(Ax) \\ &= A(\lambda x) \\ &= \lambda Ax \\ &= \lambda^2 x, \end{aligned}$$

so

$$\begin{aligned} &\lambda x = \lambda^2 x \\ \iff &\ 0 = \lambda(1-\lambda)x \\ \implies &\lambda \in \{0, 1\} \\ \implies &\lambda \geq 0. \end{aligned}$$

So since A is symmetric and all its eigenvalues are non-negative and real it follows that A is positive semi-definite.

Interpreting parameters with logarithm transformed variables

Problem 3.2:

In the following, we want to consider the interpretation of the slope coefficient in models in which dependent and/or independent variables are log-transformed. For the following derivations, use the approximation $f(z) = \exp(az) \approx 1 + az$ for $az \in \mathbb{R}$ close to zero at a suitable point.
In the following, x_0 denotes the initial level, Δx any absolute change and $\frac{\Delta x}{x_0}$ any percentage change in the regressor ($\frac{\Delta x}{x_0} = 0.01$ corresponds to a one per cent change).

(a) Given the *log-level* model

$$\log(y) = \beta_0 + \beta_1 x + \varepsilon$$

with fitted values $\hat{y}(x) = \exp(\hat{\beta}_0 + \hat{\beta}_1 x)$. Show that the following applies

$$\frac{\hat{y}(x_0 + \Delta x) - \hat{y}(x_0)}{\hat{y}(x_0)} \approx \hat{\beta}_1 \Delta x. \tag{1}$$

Interpret $\hat{\beta}_1$ for the case that $\Delta x = 1$ applies. Then transform equation (1) to $\hat{y}(x_0 + \Delta x)$ and check your interpretation using the transformation (continue to use $\Delta x = 1$).

Solution: To show: $\frac{\hat{y}(x_0+\Delta x)-\hat{y}(x_0)}{\hat{y}(x_0)} \overset{!}{\approx} \hat{\beta}_1 \Delta x$

$$\begin{aligned}\frac{\hat{y}(x_0 + \Delta x) - \hat{y}(x_0)}{\hat{y}(x_0)} &= \frac{\hat{y}(x_0 + \Delta x)}{\hat{y}(x_0)} - 1 = \frac{\exp\left[\hat{\beta}_0 + \hat{\beta}_1(x_0 + \Delta x)\right]}{\exp\left[\hat{\beta}_0 + \hat{\beta}_1 x_0\right]} - 1\\ &= \exp\left[\hat{\beta}_0 + \hat{\beta}_1(x_0 + \Delta x) - \hat{\beta}_0 - \hat{\beta}_1 x_0\right] - 1 = \exp\left[\hat{\beta}_1 \Delta x\right] - 1\end{aligned}$$

Now use the above approximation with $a := \hat{\beta}_1, z = \Delta x$:

$$\Rightarrow \frac{\hat{y}(x_0 + \Delta x) - \hat{y}(x_0)}{\hat{y}(x_0)} = \exp\left[\hat{\beta}_1 \Delta x\right] - 1 \approx 1 + \hat{\beta}_1 \Delta x - 1 = \hat{\beta}_1 \Delta x$$

If the regressor x changes by exactly one unit ($\Delta x = 1$), then the percentage change in y can be described approximately by $\hat{\beta}_1$. The slope coefficient now (approximately) indicates the effect as a percentage change in the dependent variable. Transformation of (1) results in

$$\begin{aligned}&\frac{\hat{y}(x_0 + \Delta x) - \hat{y}(x_0)}{\hat{y}(x_0)} \approx \hat{\beta}_1 \Delta x\\ \iff\ &\hat{y}(x_0 + \Delta x) \approx \hat{\beta}_1 \Delta x \hat{y}(x_0) + \hat{y}(x_0) = \hat{y}(x_0)(1 + \hat{\beta}_1 \Delta x) \qquad \mid \quad \Delta x = 1\\ \iff\ &\underbrace{\hat{y}(x_0 + 1)}_{\text{new value}} \approx \overbrace{\hat{y}(x_0)}^{\text{old value}} \cdot (1 + \underbrace{\hat{\beta}_1}_{\text{Percentage increase}}).\end{aligned}$$

(b) Given the *log-log* model

$$\log(y) = \beta_0 + \beta_1 \log(x) + \varepsilon$$

with fitted values $\hat{y}(x) = \exp(\hat{\beta}_0 + \hat{\beta}_1 \log(x))$. Show that the following applies

$$\frac{\hat{y}(x_0 + \Delta x) - \hat{y}(x_0)}{\hat{y}(x_0)} \approx \hat{\beta}_1 \frac{\Delta x}{x_0}. \tag{2}$$

Solution: Analogue to (a), we use the functional form of $\hat{y}$ for the left-hand side of the approximate equality.

$$\begin{aligned}
\frac{\hat{y}(x_0 + \Delta x) - \hat{y}(x_0)}{\hat{y}(x_0)} &= \frac{\exp\left[\hat{\beta}_0 + \hat{\beta}_1 \log(x_0 + \Delta x)\right]}{\exp\left[\hat{\beta}_0 + \hat{\beta}_1 \log(x_0)\right]} - 1 \\
&= \exp\left[\hat{\beta}_0 + \hat{\beta}_1 \log(x_0 + \Delta x) - \hat{\beta}_0 - \hat{\beta}_1 \log(x_0)\right] - 1 \\
&= \exp\left[\hat{\beta}_1 (\log(x_0 + \Delta x) - \log(x_0))\right] - 1 \\
&= \exp\left[\hat{\beta}_1 \log\left(\frac{x_0 + \Delta x}{x_0}\right)\right] - 1 \\
&= \exp\left[\hat{\beta}_1 \log\left(\frac{\Delta x}{x_0} + 1\right)\right] - 1
\end{aligned}$$

We now apply the approximation from (a) a total of two times. First 'backwards' with

$$\begin{aligned}
&a := \Delta x, z := \frac{1}{x_0} \\
\implies &\frac{\Delta x}{x_0} + 1 \approx \exp\left(\frac{\Delta x}{x_0}\right). \qquad (*)
\end{aligned}$$

We then use the approximation from (a) with

$$\begin{aligned}
&a := \hat{\beta}_1, z := \frac{\Delta x}{x_0} \\
&\exp\left[\hat{\beta}_1 \frac{\Delta x}{x_0}\right] \approx 1 + \hat{\beta}_1 \frac{\Delta x}{x_0} \qquad (**)
\end{aligned}$$

$$\begin{aligned}
\implies \frac{\hat{y}(x_0 + \Delta x) - \hat{y}(x_0)}{\hat{y}(x_0)} &= \exp\left[\hat{\beta}_1 \log\left(\frac{\Delta x}{x_0} + 1\right)\right] - 1 \\
&\overset{(*)}{\approx} \exp\left[\hat{\beta}_1 \log\left(\exp\left(\frac{\Delta x}{x_0}\right)\right)\right] - 1 \\
&= \exp\left[\hat{\beta}_1 \left(\frac{\Delta x}{x_0}\right)\right] - 1 \\
&\overset{(**)}{\approx} \hat{\beta}_1 \left(\frac{\Delta x}{x_0}\right) + 1 - 1 \\
&= \hat{\beta}_1 \left(\frac{\Delta x}{x_0}\right)
\end{aligned}$$

Interpret $\hat{\beta}_1$ for the case that $\frac{\Delta x}{x_0} = 0.01$ applies. Then solve equation (2) to $\hat{y}(x_0 + \Delta x)$ and check your interpretation (continue to use $\frac{\Delta x}{x_0} = 0.01$).

Solution: If our regressor changes by one per cent (corresponds to $\Delta x / x_0 = 0.01$), then we expect a percentage change in our dependent variable of $\frac{\hat{\beta}_1}{100}$. This results from the transformation:

$$\begin{aligned}
\hat{y}(x_0 + \Delta x) &\approx \hat{\beta}_1 \frac{\Delta x}{x_0} \hat{y}(x_0) + \hat{y}(x_0) && \Big| \quad \frac{\Delta x}{x_0} = 0.01 \\
&= \hat{y}(x_0) \cdot \left(1 + \frac{\hat{\beta}_1}{100}\right) \\
&= \hat{y}(x_0) \cdot \left(\frac{100 + \hat{\beta}_1}{100}\right)
\end{aligned}$$

(c) Given the *level-log* model

$$y = \beta_0 + \beta_1 \log(x) + \varepsilon$$

with fitted values $\hat{y}(x) = \hat{\beta}_0 + \hat{\beta}_1 \log(x)$. Show that the following applies

$$\hat{y}(x_0 + \Delta x) - \hat{y}(x_0) \approx \hat{\beta}_1 \frac{\Delta x}{x_0}. \tag{3}$$

Solution:

$$\begin{aligned}
\hat{y}(x_0 + \Delta x) - \hat{y}(x_0) &= \hat{\beta}_0 + \hat{\beta}_1 \log(x_0 + \Delta x) - \hat{\beta}_0 - \hat{\beta}_1 \log(x_0) \\
&= \hat{\beta}_1 \left[\underbrace{\log(x_0 + \Delta x) - \log(x_0)}_{= \log(\frac{\Delta x}{x_0} + 1)} \right] \\
&= \hat{\beta}_1 \left[\log\left(\frac{\Delta x}{x_0} + 1\right) \right] \\
&\overset{(*)}{\approx} \hat{\beta}_1 \left[\log\left(\exp\left(\frac{\Delta x}{x_0}\right)\right) \right] \\
&= \hat{\beta}_1 \frac{\Delta x}{x_0}
\end{aligned}$$

Interpret $\hat{\beta}_1$ for the case that $\frac{\Delta x}{x_0} = 0.01$ applies. Then solve equation (3) to $\hat{y}(x_0 + \Delta x)$ and check your interpretation (continue to use $\frac{\Delta x}{x_0} = 0.01$).

Solution: If the regressor changes by one per cent, we expect an absolute change in the dependent variable of $\frac{\hat{\beta}_1}{100}$. We can also see this by:

$$\hat{y}(x_0 + \Delta x) \approx \hat{y}(x_0) + \frac{\hat{\beta}_1}{100}$$