Economics 403
Michaelmas Test

Oliver Linton

December 9th, 2001

Selected Answers

1. We give the answers below.

   (a) This is clearly true: what is required for exact collinearity is that there exists an exact linear relationship amongst the columns of the X matrix. In such a case, the individual sample correlations between variables need not be perfect.

   (b) True, depending on the relative precisions of the two samples. Use the notation X_i, Y_j for the two different samples of size n each with the same population mean μ but population variances σ_X^2 and σ_Y^2. Let

$$\widetilde{\theta}_X = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ and } \widetilde{\theta}_Y = \frac{1}{n}\sum_{j=1}^{n} Y_j.$$

Then the question compares $\widetilde{\theta}_X$ with $\widehat{\theta} = (\widetilde{\theta}_X + \widetilde{\theta}_Y)/2$. We have

$$\text{var}(\widehat{\theta}) = \frac{1}{4}\left(\text{var}(\widetilde{\theta}_X) + \text{var}(\widetilde{\theta}_Y)\right) = \frac{\sigma_X^2 + \sigma_Y^2}{4n}$$

because the samples are mutually independent. We have

$$\text{var}(\widehat{\theta}) < \text{var}(\widetilde{\theta}_X) = \frac{\sigma_X^2}{n} \Longleftrightarrow 3\sigma_X^2 > \sigma_Y^2,$$

and likewise in the comparison between $\text{var}(\widehat{\theta})$ and $\text{var}(\widetilde{\theta}_Y)$. The result is that $\widehat{\theta}$ is best to use when $\sigma_X^2/\sigma_Y^2 \in [1/3, 3]$. Otherwise, either one of the individual samples is preferable. Note that the optimal estimator here is

$$\alpha_{opt}\widetilde{\theta}_X + (1 - \alpha_{opt})\widetilde{\theta}_Y$$

with

$$\alpha_{opt} = \frac{\sigma_Y^2}{\sigma_X^2 + \sigma_Y^2}.$$

The resulting variance is smaller than that of $\widehat{\theta}$ or that of either $\widetilde{\theta}_X$ or $\widetilde{\theta}_Y$, obviously. The lesson here is that you can always do better with more informaiton, provided you use it right, but that if you don't use it right you may end up worserer off.

(c) False. We have by the continuous mapping theorem that

$$\ln T + \ln \widehat{\theta} \xrightarrow{d} \ln \chi^2(1),$$

but the stated result is incorrect.

(d) This is debatable. We should really consider mean squared error rather than just bias or variance. In any case, adding redundant variables can actually increase bias [e.g., if you include the dependent variable on the right hand side], although in our usual text book model of exogenous or fixed regressors, adding regressors only raises variance but not bias.

(e) We would hope that the power of a test exceeds the significance level, so that there is more chance of rejecting a false model than rejecting a true model.

2. Consider the linear model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u. \tag{1}$$

If we set $z = y - x_2$, $w_1 = x_1 - x_2$, and $w_2 = x_2$, we can rewrite (1) as

$$z = \alpha_0 + \alpha_1 w_1 + \alpha_2 w_2 + v. \tag{2}$$

(a) By substituting in for z, w_1, w_2 we get that $\alpha_0 = \beta_0$, $\alpha_1 = \beta_1$, $\alpha_2 + 1 - \alpha_1 = \beta_2$, and $v = u$. This means that $\alpha_2 = \beta_1 + \beta_2 - 1$.

(b) It is clear that $\{1, w_1, w_2\}$ generate the same space as $\{1, x_1, x_2\}$ so that the residuals of the same variable on either of these sets of covariates would produce the same residuals and R^2. However, the dependent variable in the second regression is different, so what? Remember that projections are linear, so that

$$\begin{aligned} \prod(y - x_2 | span\{1, w_1, w_2\}) &= \prod(y | span\{1, w_1, w_2\}) - \prod(x_2 | span\{1, w_1, w_2\}) \\ &= \prod(y | span\{1, w_1, w_2\}) - w_2 \end{aligned}$$

because $x_2 = w_2 \in span\{1, w_1, w_2\}$. Thus the fitted value of the second regression is equal to the fitted value of the first regression minus w_2, which means that the residuals of the two regressions are the same. The parameter estimates are also identical or rather obey the exact relationship that their population counterparts predict. The R^2 depends on a comparison between the sum of squared residuals [which is the same in these two cases] and the sample variance of the dependent variable. The sample variance of the dependent variable can be different in the two cases. Indeed

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n}(z_i - \overline{z})^2 &= \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y} - (x_{2i} - \overline{x}_2))^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2 + \frac{1}{n}\sum_{i=1}^{n}(x_{2i} - \overline{x}_2)^2 - 2\frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})(x_{2i} - \overline{x}_2), \end{aligned}$$

which can be larger or smaller than $\sum_{i=1}^n (y_i - \overline{y})^2/n$ depending on the sample covariance between y, x_2.

(c) The first and last regressions correspond to the previous discussion. The second regression is also covered by this discussion at least in so far as the residuals and R^2 are concerned. The parameters there obey different correspondences, so that $\alpha_2 = \beta_2 - 1$. The third regression is somewhat different and every estimator is different in this case in an unpredictable fashion.

(d) The t-test can be tested with the final regression and testing whether $\alpha_2 = 0$ in the standard way.

3. Consider the nonlinear regression model

$$y_i = \exp(\alpha + \beta x_i) + u_i \qquad i = 1, \ldots, n \tag{1}$$

where the u_i are i.i.d. $N(0,1)$ random variables.

(a) The log-likelihood function is

$$\ell(\theta) = -\frac{n}{2}\log 2\pi - \frac{1}{2}\sum_{i=1}^n \left[y_i - \exp(\alpha + \beta x_i)\right]^2,$$

where $\theta = (\alpha, \beta)$. We have

$$\frac{\partial \ell(\theta)}{\partial \alpha} = \sum_{i=1}^n \frac{\partial \ell_i(\theta)}{\partial \alpha}, \text{ where } \frac{\partial \ell_i(\theta)}{\partial \alpha} = [y_i - \exp(\alpha + \beta x_i)]\exp(\alpha + \beta x_i)$$

$$\frac{\partial \ell(\theta)}{\partial \beta} = \sum_{i=1}^n \frac{\partial \ell_i(\theta)}{\partial \beta}, \text{ where } \frac{\partial \ell_i(\theta)}{\partial \beta} = [y_i - \exp(\alpha + \beta x_i)]\exp(\alpha + \beta x_i)x_i$$

(b) Can either calculate the second derivatives

$$\frac{\partial^2 \ell(\theta)}{\partial \alpha^2} = -\sum_{i=1}^n \exp(2\alpha + 2\beta x_i) + \sum_{i=1}^n [y_i - \exp(\alpha + \beta x_i)]\exp(\alpha + \beta x_i)$$

$$\frac{\partial^2 \ell(\theta)}{\partial \beta^2} = -\sum_{i=1}^n \exp(2\alpha + 2\beta x_i)x_i^2 + \sum_{i=1}^n [y_i - \exp(\alpha + \beta x_i)]\exp(\alpha + \beta x_i)x_i^2$$

$$\frac{\partial^2 \ell(\theta)}{\partial \alpha \partial \beta} = -\sum_{i=1}^n \exp(2\alpha + 2\beta x_i)x_i + \sum_{i=1}^n [y_i - \exp(\alpha + \beta x_i)]\exp(\alpha + \beta x_i)x_i$$

with variable step length λ

$$\theta^{[r+1]}(\lambda) = \theta^{[r]} - \lambda\left[\frac{\partial^2 \ell}{\partial\theta\partial\theta'}(\theta^{[r]})\right]^{-1}\frac{\partial \ell}{\partial \theta}(\theta^{[r]}),$$

and choose λ to max $\ell\left(\theta^{[r]}\right)$. Or use the OPG form

$$\theta^{[r+1]} = \theta^{[r]} + \left[\sum_{i=1}^n \frac{\partial \ell_i}{\partial \theta}(\theta^{[r]})\frac{\partial \ell_i}{\partial \theta'}(\theta^{[r]})\right]^{-1}\sum_{i=1}^n \frac{\partial \ell_i}{\partial \theta}(\theta^{[r]}).$$

(c) Assuming that the x_i are i.i.d. and independent of u_i, we have

$$n^{-1}\ell(\theta) \xrightarrow{P} \frac{1}{2} - \frac{1}{2}E\left[\exp(\alpha+\beta x_i) - \exp(\alpha_0+\beta_0 x_i)\right]^2$$

provided the law of large numbers can be applied. Need to say which moments. Also want uniform convergence not just pointwise. Because this is a likelihood problem, we have

$$\sqrt{n}(\widehat{\theta}-\theta_0) \xrightarrow{D} N(0, \mathcal{I}^{-1}(\theta_0)),$$

where the information matrix $\mathcal{I}(\theta_0)$ is

$$\mathcal{I}(\theta_0) = \text{var}\left[\frac{\partial \ell_i(\theta_0)}{\partial \theta}\right] = E\begin{bmatrix} \exp(2\alpha+2\beta x_i) & \exp(2\alpha+2\beta x_i) \\ \exp(2\alpha+2\beta x_i)x_i & \exp(2\alpha+2\beta x_i)x_i^2 \end{bmatrix}.$$

There are other representations.

(d) The LM test is based on

$$\frac{\partial \ell(\alpha,0)}{\partial \beta} = \sum_{i=1}^{n}\left[y_i - \exp(\alpha)\right]\exp(\alpha)x_i.$$

The second derivatives evaluated at $\beta = 0$ are

$$\frac{\partial^2 \ell(\alpha,0)}{\partial \beta^2} = -\sum_{i=1}^{n}\exp(2\alpha)x_i^2 + \sum_{i=1}^{n}[y_i - \exp(\alpha)]\exp(\alpha)x_i^2$$

We can estimate α under the null hypothesis by $\alpha^* = \ln\overline{y}$, where $\overline{y} = n^{-1}\sum_{i=1}^{n} y_i$ and then take the quadratic form

$$LM = \frac{\left[\frac{\partial \ell(\alpha^*,0)}{\partial \beta}\right]^2}{\frac{\partial^2 \ell(\alpha^*,0)}{\partial \beta^2}} = \frac{\left[\frac{1}{\sqrt{n}}\frac{\partial \ell(\alpha^*,0)}{\partial \beta}\right]^2}{\frac{1}{n}\frac{\partial^2 \ell(\alpha^*,0)}{\partial \beta^2}} \xrightarrow{D} \chi^2(1).$$

In this case, the advantage of the LM test is that one doesn't have to do nonlinear estimation, because α^* has closed form.