Methods of Economic Investigation II (EC403)
Problem Set #2

October 25, 2002

Selected Solutions

1. For unbiasedness we need that

$$E\left[\hat{\theta}(\omega_1,\omega_2)\right]=\theta \text{ for all } \theta.$$

If both $\hat{\theta}_1$and $\hat{\theta}_2$ are unbiased, then

$$E\left[\hat{\theta}(\omega_1,\omega_2)\right]=\omega_1 E(\hat{\theta}_1)+\omega_2 E(\hat{\theta}_2)=\omega_1\theta+\omega_2\theta=(\omega_1+\omega_2)\theta=\theta,$$

where the last equality holds if and only if $\omega_1+\omega_2=1$.

To be Best Unbiased $\hat{\theta}(\omega_1,\omega_2)$ must have the smallest variance among all the (unbiased) estimators that are a linear combination of $\hat{\theta}_1$ and $\hat{\theta}_2$. Without loss of generality we can now write $\omega_1=\omega$ and $\omega_2=1-\omega$. The variance of $\hat{\theta}(\omega)$ is then:[1]

$$\begin{aligned}\text{var}\left[\hat{\theta}(\omega)\right] &= \text{var}\left[\omega\hat{\theta}_1+(1-\omega)\hat{\theta}_2\right]=\omega^2\text{var}(\hat{\theta}_1)+(1-\omega)^2\text{var}(\hat{\theta}_2)\\ &\quad +2\omega(1-\omega)\text{cov}(\hat{\theta}_1,\hat{\theta}_2). \qquad (1)\end{aligned}$$

This implies that

$$\begin{aligned}\text{var}\left[\hat{\theta}(\omega)\right] &= \omega^2\sigma_1^2+(1-\omega)^2\sigma_2^2+2\omega(1-\omega)\sigma_{12}\\ &= (\sigma_1^2+\sigma_2^2-2\sigma_{12})\omega^2+(\sigma_{12}-\sigma_2^2)2\omega+\sigma_2^2.\end{aligned}$$

[1] In general if we have a linear combination of random variables $\mathbf{y}=\mathbf{a}^T\mathbf{x}$, where $\mathbf{a}$ is a vector of constants, we have

$$\begin{aligned}var(\mathbf{a}^T\mathbf{x})=\mathbf{E} &\left[(\mathbf{Cx}-\mathbf{E}(\mathbf{a}^T\mathbf{x}))^2\right]\\ &= E\left[\mathbf{a}^T(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T a\right]\\ &= \mathbf{a}^T\Sigma\mathbf{a}=\sum_{i=1}^{n}\sum_{j=1}^{n}a_i a_j\sigma_{ij},\end{aligned}$$

where the variance-covariance matrix of $\mathbf{x}$ is Σ.

The minimizer of this function with respect to ω is

$$\omega^* = \frac{\sigma_2^2 - \sigma_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}.$$

In general ω^* will depend on the relative magnitude of the variances of each estimator $(\hat{\theta}_1,\ \hat{\theta}_2)$ as well as on the covariance. Special cases are:

(a) $(\sigma_1^2 = 0$ and $\sigma_2^2 \neq 0)$ In this case $\hat{\theta}_1$ is a random variable which is different from θ only in events with probability zero ($\hat{\theta}_1$=θ almost surely). Also $\mathrm{cov}(\hat{\theta}_1, \hat{\theta}_2) \leq (\mathrm{var}(\hat{\theta}_1)\mathrm{var}(\hat{\theta}_2))^{1/2} = 0$ in this case, so $\omega^* = 1$ as expected.

(b) When the two estimators have zero covariance

$$\omega^* = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

and

$$\frac{d\omega^*}{d\sigma_2^2} = \frac{\sigma_1^2}{(\sigma_1^2 + \sigma_2^2)^2} \geqq 0$$

so that obviously as σ_2^2 increases ω^*increases as well, and $\frac{d\omega^*}{d\sigma_1^2} = -\frac{\sigma_2^2}{(\sigma_1^2+\sigma_2^2)^2} \leqq 0$ so that as σ_1^2 increases then ω^* decreases.

(c) When $\sigma_{12} = \sigma_2^2$, we set $\omega^* = 0$ so as to use only the $\hat{\theta}_2$ estimator because it is efficient in itself; in this case $\sigma_1^2 \geq \sigma_2^2 = \sigma_{12}$ because $|\mathrm{corr}(\hat{\theta}_1, \hat{\theta}_2)| \leq 1$.

2. The hypothesis for (a) is a test for the significance of the whole regression and we can use a general F-test given by:

$$F = \frac{RSS - URSS/J}{URSS/(N-K)} \sim F_{J,N-K}$$

where RSS = Restricted sum of squares= $\sum_{i=1}^{n} y_i^2$, $URSS$ = unrestricted sum of squares, J = number of restrictions.

The idea of this "loss of fit" approach is that the difference in the fitting power (as measured by R^2) between the unrestricted and the restricted regression should be stochastically small under the null hypothesis and due essentially to sampling variability, while under the alternative there should be a systematic component to the difference. In our case the null is of zero fitting power of all the regressors. We have

$$F_{4,21} = \frac{RSS - URSS/4}{URSS/21}$$

and the 5% critical value for the F distribution with 4 and 21 df is 2.84 (obviously we can test only two sided alternatives).

The hypothesis (b) can be greatly simplified by reparameterization and then using the t-test on a single parameter. Define $\gamma \equiv \beta_2 - \beta_1$. Then,

$$y_i = \beta_0 + \beta_1 x_{1i} - \beta_2 x_{2i} + \beta_1 x_{2i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i$$

$$y_i = \beta_0 + \beta_1(x_{1i} + x_{2i}) + \gamma x_{2i} + \beta_3 x_{3i} + \varepsilon_i. \quad (2)$$

Under the null $\gamma = 0$ and the corresponding statistic for the hypothesis testing is:

$$t = \frac{\hat{\gamma}}{(\text{estimated (SD) of } \hat{\gamma})},$$

where $\hat{\gamma}$ is estimated from the regression (2). Under the null hypothesis t is distributed as t_{21}: the critical value for a one side test at 5% of significance is 1.721.

The last hypothesis test is of the form $(\beta_0/\beta_1 - \beta_2/\beta_1) = 1$ against a two sided alternative. We reparameterize again such that: $\alpha \equiv \beta_0 - \beta_1 - \beta_2$, then:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i$$

becomes

$$\begin{aligned} y_i &= \beta_0 + \beta_0 - \beta_0 - \beta_1 + \beta_1 - \beta_2 + \beta_2 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i \\ y_i &= \alpha + \beta_1(x_{1i} + 1) + \beta_2(x_{2i} + 1) + \beta_3 x_{3i} + \varepsilon_i. \end{aligned}$$

Under the null hypothesis $\alpha = 0$ and the the statistic for the test is

$$t = \frac{\hat{\alpha}}{(\text{estimated (SD) of } \hat{\alpha})},$$

which is distributed as student's t with 21 df and critical values of ± 2.080 at 5% significance level.

Note:

- There isn't a unique way of reparameterization.
- We reparameterize to avoid calculation and to use routine computer outputs, these hypothesis could be tested either using the standard way of expressing restrictions on the parameters ($R\beta = r$ or $c'\beta = q$).
- In this way (due to the fact that F is the ratio of two squared random variables) we can test only two sided hypothesis.

3. We know that if a constant is included among the regressors the three versions of the R^2 lead to the same values, but if not: the first one ($R_1^2 = 1 - \frac{RSS}{TSS}$) can be negative, while the third one ($R_3^2 = \frac{ESS}{TSS}$) can be greater than one; the second one ($R_2^2 = [\text{corr}(y, \hat{y})]^2$) always lies between 0 and 1 (use the Cauchy-Schwarz inequality).

A simple example can be the one in which I have only two observations (for simplicity) that are respectively a "little above" and a "little below" their mean, such that $\sum(y_i - \bar{y})^2 \approx 0^+$, the fitted values are computed without a constant term such that in our case the regression can be represented in a Cartesian plane with a line through the origin, then $\sum(y_i - \hat{y}_i)^2$ is not unbounded. Hence:

$$R_1^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \approx 1 - \frac{K}{0^+} \approx -\infty$$

$$R_3^2 = \frac{ESS}{TSS} = \frac{\sum(\hat{y}_i - \bar{y}_i)^2}{\sum(y_i - \bar{y})^2} \approx \frac{K}{0^+} \approx +\infty.$$

4. The student's t distribution is given by the ratio between a standard normal r.v. and the square root of a chi-squared divided by its degrees of freedom, i.e.:

$$t = \frac{U}{\sqrt{X/n}} \sim t_{(n)}$$

where $U \sim N(0,1)$, $X \sim \chi^2_{(n)}$ and U and X are independent random variables. At the same time the F distribution is given by the ratio between two independent chi-squared r.v. divided by their degrees of freedom, i.e.:

$$F = \frac{X_1/n_1}{X_2/n_2} \sim F\left[n_1, n_2\right].$$

A simple manipulation of the t r.v. leads to a precise class of F r.v.:

$$t^2 = \frac{(U)^2 \equiv \chi_{(1)}}{X_{(n)}/n} = \frac{X_{(1)}}{X_{(n)}/n} \sim F\left[1, n\right]$$

From this relation we can derive straightforwardly another one that employs the R^2 as well. Suppose we have a multiple regression model $y = X\beta + \varepsilon$ and then compute the R-squared defined for this specification as R^2_X. then add another regressor z , the change in the R-squared will be: $R^2_{Xz} = R^2_X + (1 - R^2_X) * r^{*2}_{yz}$, where r^{*2}_{yz} is the square of the partial correlation between y and z. But

$$r^{*2}_{yz} = \frac{t_z^2}{t_z^2 + (n-K)}.$$

Then rearranging

$$t^2_{z(n-k)} = \frac{(R^2_{Xz} - R^2_X)}{(1 - R^2_{Xz})/(n-k)} = F_{[1,n-k]}.$$

The relation between the F-test and the t-test is explored in the second part of the exercise. We have three hypotheses to test: the first one is a joint hypothesis about the values of both the coefficient while the remaining two are about the value of a single coefficient. We have:

- $u_i \sim N(0,1)$ and $\text{cov}(u_i, u_j) = 0 \ \forall \ i \neq j$
- $K = 3$

- $\frac{X'X}{n} = I_3$
- An intercept is included

The second hypothesis $\text{H}_0 : \beta_2 = 0$ vs. H_A : not H_0, implies that

$$\hat{\beta}_2 \sim N(\beta_2, \frac{1}{n})$$

under the null. The acceptance region is then

$$|\hat{\beta}_2| \leq z_{\alpha/2} * \frac{1}{\sqrt{n}}.$$

The same applies to the third hypothesis $\text{H}_0 : \beta_3 = 0$ vs. H_A : not H_0, for which $\widehat{\beta}_3 \sim N(\beta_3, \frac{1}{n})$ under the null. The acceptance region is then:

$$|\hat{\beta}_3| \leq z_{\alpha/2} * \frac{1}{\sqrt{n}}.$$

For the first hypothesis $\text{H}_0 : \beta_2 = \beta_3 = 0$ vs. H_A : not H_0 we use the Wald test: $\text{H}_0 : c(\theta) = q$ vs $\text{H}_A : c(\theta) \neq q$,

$$W = \left[c(\hat{\theta}) - q\right]' \left[\text{var}(c(\hat{\theta}))\right]^{-1} \left[c(\hat{\theta}) - q\right],$$

which can be expressed for our linear restriction as:

$$c(\hat{\theta}) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} \; ; \; q = [0, 0] .$$

Therefore,

$$W = \begin{bmatrix} \hat{\beta}_2 & \hat{\beta}_3 \end{bmatrix} \begin{bmatrix} \frac{1}{n} & 0 \\ 0 & \frac{1}{n} \end{bmatrix}^{-1} \begin{bmatrix} \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = n(\hat{\beta}_2^2 + \hat{\beta}_3^2) \sim \chi^2_{(2)},$$

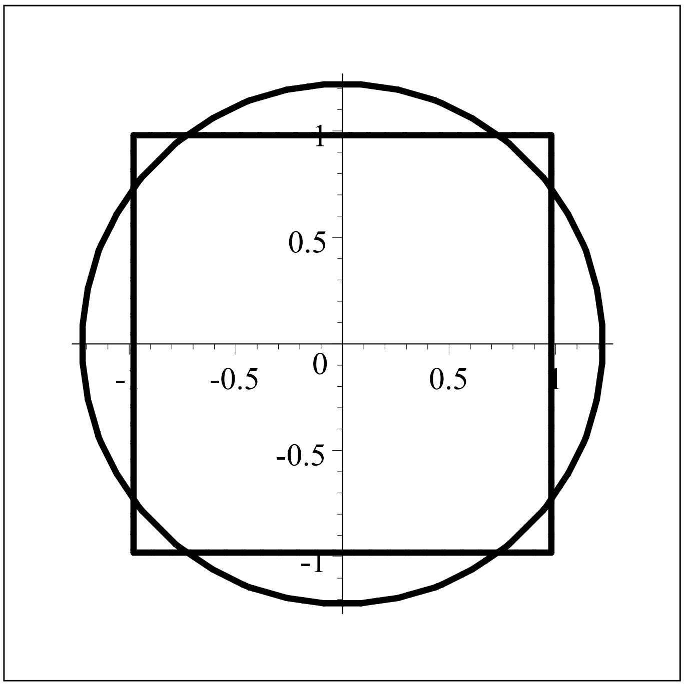
and you reject if

$$(\hat{\beta}_2^2 + \hat{\beta}_3^2) \geq \frac{\chi^2_{(2),\alpha}}{n}.$$

The critical values for $\alpha = 0.05$ and $n = 4$ are

- $|\hat{\beta}_2| \leq z_{0.05/2} * \frac{1}{2} = 0.98$
- $|\hat{\beta}_3| \leq z_{0.05/2} * \frac{1}{2} = 0.98$
- $(\hat{\beta}_2^2 + \hat{\beta}_3^2) \leq \frac{\chi^2_{(2),0.05}}{4} \approx 1.5$

We can give a nice graphical representation of the joint analysis of these tests: $\hat{\beta}_2^2 + \hat{\beta}_3^2 \leq 1.5$, $|\hat{\beta}_2| \leq 0.98$, $|\hat{\beta}_3| \leq 0.98$



We can then conclude that:

(a) It is possible to accept both H_0^b and H_0^c but to reject H_0^a [area inside the square but outside the circle].

(b) It is not possible (with orthogonal regressors) to accept H_0^a and to reject both H_0^b and H_0^c.

(c) If you had non-orthogonal regressors, then the answer to (b) could be changed. In that case we have an ellipse rather than a circle and it could extend arbitrarily far in the 45 degree direction.

(d) The probability content of the square region under H_0^a is $0.95^2 = 0.9025$. Therefore the test that combines b and c has rejection rate nearly 10%. Should set wider critical values to obtain valid test at 5% level.

5. (a) is false, let's see why. Let $X \sim N(\mu, 1)$ then for an i.i.d sample

$$\bar{X} = \left(\frac{1}{n} \sum_{i=1}^{n} X_i \right) \sim N(\mu, \frac{1}{n})$$

and $Z = \sqrt{n}\left(\bar{X} - \mu\right) \sim N(0,1)$. Therefore, $\Pr(Z < 1.645) = 0.95$ and $\Pr(|Z| < 1.96) = 0.95$. These are two 95% confidence intervals where none of the two strictly contains the other:

- The first one is given by $(\bar{X} - \frac{1.645}{\sqrt{n}}, +\infty)$
- The second one by $(\bar{X} - \frac{1.96}{\sqrt{n}}, \bar{X} + \frac{1.96}{\sqrt{n}})$

(b) is, as already seen in exercise 3, true if the regression equation contains an intercept otherwise is false for R_1^2 and R_3^2 while is always true, with or without an intercept, for R_2^2.

(c) the answer is false. The starting regression equation is:

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

The estimated coefficient of this univariate model are as usual:

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \text{ and } \hat{\beta} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}.$$

Then changing only the units of both the weight and the height variables: $y_i^* = \theta y_i$ and $x_i^* = \rho x_i$ we have

$$\hat{\alpha}^* = \bar{y}^* - \hat{\beta}^*\bar{x}^* \text{ and } \hat{\beta}^* = \frac{\sum(x_i^* - \bar{x}^*)(y_i^* - \bar{y}^*)}{\sum(x_i^* - \bar{x}^*)^2}.$$

Substituting we get

$$\hat{\beta}^* = \frac{\sum(\rho x_i - \rho\bar{x})(\theta y_i - \theta\bar{y})}{\sum(\rho x_i - \rho\bar{x})^2} = \frac{\rho * \theta}{\rho^2}\frac{\sum(x_i^* - \bar{x}^*)(y_i^* - \bar{y}^*)}{\sum(x_i^* - \bar{x}^*)^2} = \frac{\theta}{\rho}\frac{\sum(x_i^* - \bar{x}^*)(y_i^* - \bar{y}^*)}{\sum(x_i^* - \bar{x}^*)^2} = \frac{\theta}{\rho}\hat{\beta}$$

and

$$\hat{\alpha}^* = \theta\bar{y} - \frac{\theta}{\rho}\hat{\beta}\rho\bar{x} = \theta(\bar{y} - \hat{\beta}\bar{x}) = \theta\hat{\alpha},$$

which are different in the two regressions.

For the point (d) the answer is that it is usually true even if not always true; let's see why. The Model is

$$y = X\beta + \varepsilon,$$

where all the variables have been demeaned and the usual assumptions applies. The value of the single coefficient, say $\hat{\beta}_{(j)}$, is

$$\hat{\beta}_{(j)} = (x_j' M_j x_j)^{-1} x_j M_j y,$$

where x_j is the $j-th$ regressor and $M_j = I_n - X_{(j)}(X_{(j)}'X_{(j)})^{-1}X_{(j)}'$ is the "residual maker" matrix (less the $j-th$ regressor). The variance for this estimator is

$$\text{var}(\hat{\beta}_{(j)}) = \sigma^2(x_j' M_j x_j)^{-1} = \sigma^2\left[x_j'x_j - x_j'X_{(j)}(X_{(j)}'X_{(j)})^{-1}X_{(j)}'x_j\right]^{-1}$$

$$= \sigma^2\left\{x_j'x_j\left[1 - \frac{x_j'X_{(j)}(X_{(j)}'X_{(j)})^{-1}X_{(j)}'x_j}{x_j'x_j}\right]\right\}^{-1},$$

but

$$\frac{x_j'X_{(j)}(X_{(j)}'X_{(j)})^{-1}X_{(j)}'x_j}{x_j'x_j} = \frac{x_j'P_{(j)}'P_{(j)}x_j}{x_j'x_j} = R^2_{x_j,X_{(j)}} = \text{ the R-squared of the regr. of } x_j \text{ on } X_{(j)}.$$

Therefore,

$$\text{var}(\hat{\beta}_{(j)}) = \sigma^2 \left[x_j' x_j (1 - R_j^2)\right]^{-1} = \frac{\sigma^2}{x_j' x_j (1 - R_j^2)}.$$

If the variables are highly correlated $R_j^2 \longrightarrow 1$ (while ss_j remains bounded) and hence $\text{var}(\hat{\beta}_{(j)}) \longrightarrow \infty$

While at the same time the $F - test$ will remain significant. We can see this exercise as a counter-example of the exercise 4 about multicolllinearity.