AS.180.633: Econometrics **Spring 2020**

Homework 7: Suggested Solutions

Instructor: Yingyao Hu *By:* Tong Zhou

8.18

Suppose you have two independent samples

$$y_{1i} = \boldsymbol{x}_{1i}'\boldsymbol{\beta}_1 + e_{1i}$$

and

$$y_{2i} = \boldsymbol{x}_{2i}'\boldsymbol{\beta}_2 + e_{2i}$$

both of sample size n, and both $\boldsymbol{x}_{1i}$ and $\boldsymbol{x}_{2i}$ are $k \times 1$. You estimate $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ by OLS on each sample, $\widehat{\boldsymbol{\beta}}_1$ and $\widehat{\boldsymbol{\beta}}_2$, say, with asymptotic covariance matrix estimators $\widehat{V}_{\widehat{\beta}_1}$ and $\widehat{V}_{\widehat{\beta}_2}$. Consider efficient minimum distance estimation under the restriction $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$.

(a) Find the estimator $\widetilde{\boldsymbol{\beta}}$ of $\boldsymbol{\beta} = \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$.

(b) Find the asymptotic distribution of $\widetilde{\boldsymbol{\beta}}$.

(c) How would you approach the problem if the sample sizes are different, say n_1 and n_2?

Proof. We shall start with part (c), because part (a) and (b) are special cases.

(c) Mimicking the objective function in 8.19, when n_1 and n_2 are different,

$$\begin{aligned} J(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) &= \frac{1}{2}\begin{pmatrix} n_1\left(\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1\right) \\ n_2\left(\widehat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2\right)\end{pmatrix}' \begin{pmatrix} \boldsymbol{V}_1^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{V}_2^{-1}\end{pmatrix} \begin{pmatrix} \widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1 \\ \widehat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2\end{pmatrix} \\ &= \frac{1}{2} n_1 \left(\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1\right)' \boldsymbol{V}_1^{-1}\left(\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1\right) + \frac{1}{2} n_2 \left(\widehat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2\right)' \boldsymbol{V}_2^{-1}\left(\widehat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2\right) \end{aligned}$$

and the constraint is

$$\boldsymbol{R}'\boldsymbol{\beta} = (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2) = \boldsymbol{0}$$

The Lagrangian function is

$$\mathscr{L}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\lambda}) = J(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) + \boldsymbol{\lambda}'(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)$$

The FOC is:

$$\begin{cases} \dfrac{\partial \mathscr{L}}{\partial \boldsymbol{\beta}_1} = -n_1 \boldsymbol{V}_1^{-1}\widehat{\boldsymbol{\beta}}_1 + n_1 \boldsymbol{V}_1^{-1}\widetilde{\boldsymbol{\beta}}_1 + \widetilde{\boldsymbol{\lambda}} = \mathbf{0} \\ \dfrac{\partial \mathscr{L}}{\partial \boldsymbol{\beta}_2} = -n_2 \boldsymbol{V}_2^{-1}\widehat{\boldsymbol{\beta}}_2 + n_2 \boldsymbol{V}_2^{-1}\widetilde{\boldsymbol{\beta}}_2 - \widetilde{\boldsymbol{\lambda}} = \mathbf{0} \\ \dfrac{\partial \mathscr{L}}{\partial \boldsymbol{\lambda}} = \widetilde{\boldsymbol{\beta}}_1 - \widetilde{\boldsymbol{\beta}}_2 = \mathbf{0} \end{cases}$$

From it, we have

$$\begin{aligned} \widetilde{\boldsymbol{\beta}} &= \left(n_1 \boldsymbol{V}_1^{-1} + n_2 \boldsymbol{V}_2^{-1}\right)^{-1}\left(n_1 \boldsymbol{V}_1^{-1}\widehat{\boldsymbol{\beta}}_1 + n_2 \boldsymbol{V}_2^{-1}\widehat{\boldsymbol{\beta}}_2\right) \\ &= \left[\frac{n_1}{n_2}\boldsymbol{V}_1^{-1} + \boldsymbol{V}_2^{-1}\right]^{-1}\left[\frac{n_1}{n_2}\boldsymbol{V}_1^{-1}\widehat{\boldsymbol{\beta}}_1 + \boldsymbol{V}_2^{-1}\widehat{\boldsymbol{\beta}}_2\right] \end{aligned}$$

Therefore, the asymptotic behavior of $\widetilde{\boldsymbol{\beta}}$ totally depends on the convergence rate $\frac{n_1}{n_2}$.

Case 1. $n_1 = n_2 = n$

This case reduces to part (a). It is straightforward that

$$\widetilde{\boldsymbol{\beta}} = \left(\boldsymbol{V}_1^{-1} + \boldsymbol{V}_2^{-1}\right)^{-1}\left(\boldsymbol{V}_1^{-1}\widehat{\boldsymbol{\beta}}_1 + \boldsymbol{V}_2^{-1}\widehat{\boldsymbol{\beta}}_2\right)$$

and by

$$\sqrt{n}\,(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \left(\boldsymbol{V}_1^{-1} + \boldsymbol{V}_2^{-1}\right)^{-1}\boldsymbol{V}_1^{-1}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}\right) + \left(\boldsymbol{V}_1^{-1} + \boldsymbol{V}_2^{-1}\right)^{-1}\boldsymbol{V}_2^{-1}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}\right)$$

Define $\Omega_1 = \left(\boldsymbol{V}_1^{-1} + \boldsymbol{V}_2^{-1}\right)^{-1}\boldsymbol{V}_1^{-1}$ and $\Omega_2 = \left(\boldsymbol{V}_1^{-1} + \boldsymbol{V}_2^{-1}\right)^{-1}\boldsymbol{V}_2^{-1}$

Then we have

$$\Omega_1\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}\right) \xrightarrow{\mathsf{d}} \mathcal{N}\left(\mathbf{0}, \left(\boldsymbol{V}_1^{-1} + \boldsymbol{V}_2^{-1}\right)^{-1}\boldsymbol{V}_1^{-1}\left(\boldsymbol{V}_1^{-1} + \boldsymbol{V}_2^{-1}\right)^{-1}\right)$$

and

$$\Omega_2\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}\right) \xrightarrow{\mathsf{d}} \mathcal{N}\left(\mathbf{0}, \left(\boldsymbol{V}_1^{-1} + \boldsymbol{V}_2^{-1}\right)^{-1}\boldsymbol{V}_2^{-1}\left(\boldsymbol{V}_1^{-1} + \boldsymbol{V}_2^{-1}\right)^{-1}\right)$$

Since they are independent,

$$\sqrt{n}\,(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\mathsf{d}} \mathcal{N}\left(\mathbf{0}, \left(\boldsymbol{V}_1^{-1} + \boldsymbol{V}_2^{-1}\right)^{-1}\right)$$

Case 2: $\frac{n_1}{n_2} \to 0$.

In this case, $\widetilde{\boldsymbol{\beta}} \xrightarrow{\mathsf{P}} \widehat{\boldsymbol{\beta}}_2$ and

$$\sqrt{n_1 + n_2}\,(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\mathsf{P}} \mathcal{N}\left(\mathbf{0}, \boldsymbol{V}_2\right)$$

Case 3: $\frac{n_1}{n_2} \to \infty$.

Similarly,

$$\sqrt{n_1+n_2}\left(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}\right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{V}_1)$$

Case 4: $\frac{n_1}{n_2} \to c > 0$.

It is easy to show that

$$\sqrt{n_1+n_2}\left(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, (1+c)\left(c\boldsymbol{V}_1^{-1}+\boldsymbol{V}_2^{-1}\right)^{-1}\right)$$

It is easily verified that when $c=1$, it reduces to **Case 1**. □

8.19

As in Exercise 7.28 and 3.26, sue the CPS dataset and subsample of white male Hispanics,

(a) Estimate the regression

$$\begin{aligned}\widehat{\log(\text{wage})} = {} & \beta_1\text{education} + \beta_2\text{experience} + \beta_3\text{experience}^2/100 + \beta_4\text{married}_1 \\ & + \beta_5\text{married}_2 + \beta_6\text{married}_3 + \beta_7\text{widowed} + \beta_8\text{divorced} + \beta_9\text{separated} + \beta_{10}\end{aligned}$$

(b) Estimate the equation using constrained least-squares, imposing the constraints $\beta_4 = \beta_7$ and $\beta_8 = \beta_9$, and report the estimates and standard errors.

(c) Estimate the equation using efficient minimum distance imposing the same constraints. Report the estimates and standard errors.

(d) Under what constraint on the coefficients is the wage equation non-decreasing in experience for experience up to 50?

(e) Estimate the equation imposing $\beta_4 = \beta_7$, $\beta_8 = \beta_9$ and the inequality from part (d).

(a) (b) (c) See Table 1, Table 2 and Table 3

(d) The constraint is $\beta_2 \geq 0$ and $\beta_2 + \beta_3 \geq 0$.

(e) Observe that $\widehat{\beta}_{2,\text{emd}} > 0$ and $\widehat{\beta}_{2,\text{emd}} + \widehat{\beta}_{3,\text{emd}} < 0$ in Table 3 , so the inequality constraint is equivalent to imposing a constraint $\beta_2 + \beta_3 = 0$. See results in Table 4

Note that R codes are on page 13.

Table 1: OLS Estimates and s.e.

	Estimates	se
education	0.089	0.003
experience	0.030	0.003
experience2	-0.037	0.006
married 1	0.181	0.025
married 2	-0.480	0.033
married 3	-0.040	0.056
widow	0.236	0.173
divorced	0.074	0.045
sep	0.016	0.053
intercept	1.192	0.046

Table 2: CLS Estimates and s.e.

	Estimates	se
education	0.089	0.003
experience	0.030	0.003
experience 2	-0.037	0.005
married 1	0.180	0.024
married 2	-0.479	0.572
married 3	-0.040	0.057
widow	0.180	0.024
divorced	0.055	0.038
sep	0.055	0.038
intercept	1.189	0.045

Table 3: EMD estimates and s.e.

	Estimates	se
education	0.089	0.003
experience	0.030	0.003
experience 2	-0.037	0.006
married 1	0.180	0.025
married 2	-0.480	0.033
married 3	-0.040	0.056
widow	0.180	0.025
divorced	0.050	0.038
sep	0.050	0.038
intercept	1.188	0.046

Table 4: EMD Estimates and s.e. Under Inequality Constraint

	Estimates	se
education	0.090	0.003
experience	0.024	0.002
experience 2	-0.024	0.002
married 1	0.190	0.025
married 2	-0.487	0.033
married 3	-0.035	0.056
widow	0.190	0.025
divorced	0.063	0.038
sep	0.063	0.038
intercept	1.216	0.044

9.18

The observed data is $\{y_i, \boldsymbol{x}_i, \boldsymbol{z}_i\} \in \mathbb{R} \times \mathbb{R}^k \times \mathbb{R}^l, k > 1$ and $l > 1, i = 1, \cdots, n$. An econometrician first estimates

$$y_i = \boldsymbol{x}_i' \widehat{\boldsymbol{\beta}} + \widehat{e}_i$$

by least squares. The econometrician next regresses the residual $\widehat{e}_i$ on $\boldsymbol{z}_i$, which can be written as

$$\widehat{e}_i = \boldsymbol{z}_i' \widetilde{\boldsymbol{\gamma}} + \widetilde{u}_i.$$

(a) Define the population parameter $\boldsymbol{\gamma}$ being estimated in this second regression.

(b) Find the probability limit of $\widetilde{\boldsymbol{\gamma}}$.

(c) Suppose the econometrician constructs a Wald statistic $\mathcal{W}_n$ for $\mathbb{H}_0 : \boldsymbol{\gamma} = \mathbf{0}$ from the second regression, ignoring the regression. Write down the formula for $\mathcal{W}_n$.

(d) Assuming $\mathbb{E}[\boldsymbol{z}_i \boldsymbol{x}_i'] = \mathbf{0}$, find the asymptotic distribution for $\mathcal{W}_n$ under $\mathbb{H}_0 : \boldsymbol{\gamma} = \mathbf{0}$.

(e) If $\mathbb{E}[\boldsymbol{z}_i \boldsymbol{x}_i'] \neq 0$ will your answer to (d) change?

Proof. **(a)**

$\boldsymbol{\gamma}$ is the linear projection coefficient of e_i on $\boldsymbol{z}_i$, i.e.

$$e_i = \boldsymbol{z}_i' \boldsymbol{\gamma} + u_i \tag{1}$$

and

$$\boldsymbol{\gamma} = \left(\mathbb{E}\left[\boldsymbol{z}_i \boldsymbol{z}_i'\right]\right)^{-1} \mathbb{E}\left[\boldsymbol{z}_i e_i\right]$$

(b)

By $\widehat{e}_i = y_i - \boldsymbol{x}_i'\widehat{\boldsymbol{\beta}} = e_i + \boldsymbol{x}_i'\left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\right)$,

the OLS estimator

$$\begin{aligned}
\widetilde{\boldsymbol{\gamma}} &= \left(\sum_{i=1}^{n} \boldsymbol{z}_i \boldsymbol{z}_i'\right)^{-1} \left(\sum_{i=1}^{n} \boldsymbol{z}_i \widehat{e}_i\right) \\
&= \left(\frac{1}{n}\sum_i \boldsymbol{z}_i \boldsymbol{z}_i'\right)^{-1} \left(\frac{1}{n}\sum_i \boldsymbol{z}_i e_i\right) + \left(\frac{1}{n}\sum_i \boldsymbol{z}_i \boldsymbol{z}_i'\right)^{-1} \left(\frac{1}{n}\sum_i \boldsymbol{z}_i \boldsymbol{x}_i'\right)\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \qquad (2) \\
&\xrightarrow{\mathsf{P}} \left(\mathbb{E}\left[\boldsymbol{z}_i \boldsymbol{z}_i'\right]\right)^{-1} \mathbb{E}[\boldsymbol{z}_i e_i] \equiv \boldsymbol{\gamma}
\end{aligned}$$

That is,

$$\widetilde{\boldsymbol{\gamma}} \xrightarrow{\mathsf{P}} \boldsymbol{\gamma}$$

(c)

Suppose

$$\sqrt{n}\left(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}\right) \xrightarrow{\mathsf{d}} \mathcal{N}\left(\mathbf{0}, \boldsymbol{V}_{\boldsymbol{\gamma}}\right).$$

Then under $\mathbb{H}_0 : \boldsymbol{\gamma} = \mathbf{0}$, the Wald statistic

$$\mathcal{W}_n = n\widehat{\boldsymbol{\gamma}}'\widehat{\boldsymbol{V}}_{\boldsymbol{\gamma}}^{-1}\widehat{\boldsymbol{\gamma}} \sim \chi_l^2$$

(d)

Continuing on **(b)**, plug Eq. (1) into Eq. (2)

$$\widetilde{\boldsymbol{\gamma}} = \left(\frac{1}{n}\sum_i \boldsymbol{z}_i \boldsymbol{z}_i'\right)^{-1} \left(\frac{1}{n}\sum_i \boldsymbol{z}_i \boldsymbol{x}_i'\right)\underbrace{\left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\right)}_{=\boldsymbol{o}_p(1)} + \boldsymbol{\gamma} + \left(\frac{1}{n}\sum_i \boldsymbol{z}_i \boldsymbol{z}_i'\right)^{-1} \left(\frac{1}{n}\sum_i \boldsymbol{z}_i u_i\right)$$

and $\mathbb{E}[\boldsymbol{z}_i \boldsymbol{x}_i'] = \mathbf{0}$ implies $\frac{1}{n}\sum_i \boldsymbol{z}_i \boldsymbol{x}_i' \xrightarrow{\mathsf{P}} \mathbf{0}$, therefore

$$\sqrt{n}\left(\widetilde{\boldsymbol{\gamma}} - \boldsymbol{\gamma}\right) = \left(\frac{1}{n}\sum_i \boldsymbol{z}_i \boldsymbol{z}_i'\right)^{-1} \underbrace{\left(\frac{1}{n}\sum_i \boldsymbol{z}_i \boldsymbol{x}_i'\right)}_{=\boldsymbol{o}_p(1)} \underbrace{\sqrt{n}\left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\right)}_{=\boldsymbol{O}_p(1)} + \left(\frac{1}{n}\sum_i \boldsymbol{z}_i \boldsymbol{z}_i'\right)^{-1} \left(\frac{1}{\sqrt{n}}\sum_i \boldsymbol{z}_i u_i\right)$$

By WLLN and CLT,

$$\frac{1}{n}\sum_i \boldsymbol{z}_i\boldsymbol{z}_i' \xrightarrow{\text{P}} \mathbb{E}[\boldsymbol{z}_i\boldsymbol{z}_i'] \equiv \boldsymbol{Q}_{zz}$$
$$\frac{1}{\sqrt{n}}\sum_i \boldsymbol{z}_i u_i \xrightarrow{\text{d}} \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{\Omega}_u\right)$$

where $\Omega_u = \mathbb{E}[\boldsymbol{z}_i\boldsymbol{z}_i' u_i^2]$.

Therefore

$$\sqrt{n}\left(\widetilde{\boldsymbol{\gamma}} - \boldsymbol{\gamma}\right) \xrightarrow{\text{d}} \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{Q}_{zz}^{-1}\boldsymbol{\Omega}_u\boldsymbol{Q}_{zz}^{-1}\right)$$

and the Wald statistic under $\mathbb{H}_0 : \boldsymbol{\gamma} = \boldsymbol{0}$ is

$$\mathcal{W}_n = n\widetilde{\boldsymbol{\gamma}}'\widehat{\boldsymbol{Q}}_{zz}\widehat{\boldsymbol{\Omega}}_u^{-1}\widehat{\boldsymbol{Q}}_{zz}\widehat{\boldsymbol{\gamma}}$$

where

$$\widehat{\boldsymbol{Q}}_{zz} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{z}_i\boldsymbol{z}_i'$$
$$\widehat{\boldsymbol{\Omega}}_u = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{z}_i\boldsymbol{z}_i'\widetilde{u}_i^2$$

(e)

The answer does change, because

$$\sqrt{n}\left(\widetilde{\boldsymbol{\gamma}} - \boldsymbol{\gamma}\right) = \underbrace{\left(\frac{1}{n}\sum_i \boldsymbol{z}_i\boldsymbol{z}_i'\right)^{-1}\underbrace{\left(\frac{1}{n}\sum_i \boldsymbol{z}_i\boldsymbol{x}_i'\right)}_{=\boldsymbol{O}_p(1)}\underbrace{\sqrt{n}\left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\right)}_{=\boldsymbol{O}_p(1)}}_{\xrightarrow{\text{d}}\mathcal{N}\left(\boldsymbol{0}, Q_{zz}^{-1}Q_{zx}\boldsymbol{V}_{\boldsymbol{\beta}}Q_{zx}'Q_{zz}^{-1}\right)} + \underbrace{\left(\frac{1}{n}\sum_i \boldsymbol{z}_i\boldsymbol{z}_i'\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_i \boldsymbol{z}_i u_i\right)}_{\xrightarrow{\text{d}}\mathcal{N}\left(\boldsymbol{0}, Q_{zz}^{-1}\boldsymbol{\Omega}_u Q_{zz}^{-1}\right)}$$

That is, the first term on the right hand side is $O_p(1)$ and will not vanish when $n \to \infty$. As a result, the asymptotic distribution of $\widetilde{\boldsymbol{\gamma}}$ will be a mixture of two asymptotic normal distributions. □

9.20

You are reading a paper, and it reports the results from two nested OLS regressions:

$$y_i = \boldsymbol{x}_{1i}'\widetilde{\boldsymbol{\beta}}_1 + \widetilde{e}_i$$
$$y_i = \boldsymbol{x}_{1i}'\widehat{\boldsymbol{\beta}}_1 + x_{2i}'\widehat{\boldsymbol{\beta}}_2 + \widehat{e}_i.$$

Some summary statistics are reported:

Short Regression	Long Regression
$R^2 = .20$	$R^2 = .26$
$\sum_{i=1}^n \widetilde{e}_i^2 = 106$	$\sum_{i=1}^n \widehat{e}_i^2 = 100$
# of coefficients = 5	# of coefficients = 8
$n = 50$	$n = 50$

Proof. It depends on how you perceive the sample size $n = 50$.

If you believe $n = 50$ is a small sample, then based on the given information, there is no hope of conducting a hypothesis testing procedure, even if the homoskedasticity assumption is made. However, under the normality condition, F test can be used to do the testing.

Recall that under the normality assumption, we have two independent χ^2-distributed statistics:

$$\frac{(n-k)s^2}{\sigma^2} \sim \chi^2_{n-k}$$

and

$$\left(\boldsymbol{R}'\widehat{\boldsymbol{\beta}} - \boldsymbol{r}\right)' \left(\sigma^2 \boldsymbol{R}' \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{R}\right)^{-1} \left(\boldsymbol{R}'\widehat{\boldsymbol{\beta}} - \boldsymbol{r}\right) \sim \chi^2_q$$

Therefore, their ratio divided by respective degrees of freedom follows a F-distribution (Equation 9.12 of Hansen's book)

$$\frac{\left(\boldsymbol{R}'\widehat{\boldsymbol{\beta}} - \boldsymbol{r}\right)' \left(\sigma^2 \boldsymbol{R}' \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{R}\right)^{-1} \left(\boldsymbol{R}'\widehat{\boldsymbol{\beta}} - \boldsymbol{r}\right)/q}{s^2/\sigma^2} = \frac{(\widetilde{\sigma}^2 - \widehat{\sigma}^2)/q}{\widehat{\sigma}^2/(n-k)} = \frac{\left(R_L^2 - R_S^2\right)/q}{\left(1 - R_L^2\right)/(n-k)} \sim F_{(q,n-k)}$$

where R_L^2 and R_S^2 are R^2's of the long regression and the short regression respectively, and $n = 50, k = 8$ and $q = 3$.

If you believe $n = 50$ is a large sample, then only the homoskedasticity condition is sufficient for hypothesis testing. By **Theorem 9.6**, the above F-statistics

$$\frac{\left(\tilde{\sigma}^2 - \hat{\sigma}^2\right)/q}{\hat{\sigma}^2/(n-k)} \xrightarrow{d} \frac{\chi_q^2}{q},$$

where the χ_q^2 distribution does not depend on the normality assumption.

Both tests give the same result that the null hypothesis $\boldsymbol{\beta}_2 = \mathbf{0}$ cannot be rejected. □

9.24

Do a Monte Carlo simulation. Take the model

$$y_i = \alpha + x_i\beta + e_i$$
$$\mathbb{E}[x_i e_i] = 0$$

where the parameter of interest is $\theta = \exp(\beta)$. Your data generating process (DGP) for the simulation is: x_i is $U[0,1]$, e_i is independent of x_i and $\mathcal{N}(0,1)$, $n = 50$. Set $\alpha = 0$ and $\beta = 1$. Generate $B = 1000$ independent samples with α. On each, estimate the regression by least-squares, calculate the covariance matrix using a standard (heteroskedasticity-robust) formula, and similarly estimate θ and its standard error. For each replication, store $\hat{\beta}, \hat{\theta}$, $t_\beta = \left(\hat{\beta} - \beta\right)/s\left(\hat{\beta}\right)$, and $t_\theta = \left(\hat{\theta} - \theta\right)/s\left(\hat{\theta}\right)$.

(a) Does the value of α matter? Explain why the described statistics are **invariant** to α and thus setting $\alpha = 0$ is irrelevant.

(b) From the 1000 replications estimate $\mathbb{E}\left[\hat{\beta}\right]$ and $\mathbb{E}\left[\hat{\theta}\right]$. Discuss if you see evidence if either estimator is biased or unbiased.

(c) From the 1000 replications estimate $\mathbb{P}\left[t_\beta > 1.645\right]$ and $\mathbb{P}\left[t_\theta > 1.645\right]$. What does asymptotic theory predict these probabilities should be in large samples? What do your simulation results indicate?

Proof. Note: R codes is on page 15

(a)

Since the two ratios do not rely on $\hat{\alpha}$, setting $\alpha = 0$ is irrelevant.

(b)

From the simulation (codes are provided below), the bias for $\widehat{\beta}$ and $\widehat{\theta}$ are

$$\text{Bias}\left(\widehat{\beta}\right) = \mathbb{E}\left[\widehat{\beta}\right] - \beta = 0.0296$$
$$\text{Bias}\left(\widehat{\theta}\right) = \mathbb{E}\left[\widehat{\theta}\right] - \theta = 0.46$$

So $\widehat{\theta}$ is biased. Though the continuous mapping theorem implies $\widehat{\theta} \xrightarrow{\text{P}} \theta$, $n = 50$ is a small sample size for a nonlinear transformation of β. Also, this upward bias can be justified by the Jensen's Inequality:

$$\mathbb{E}\left[\widehat{\theta}\right] = \mathbb{E}\left[e^{\widehat{\beta}}\right] > e^{\mathbb{E}\left[\widehat{\beta}\right]} = \theta$$

(c)

From the 1000 replications,

$$\mathbb{P}\left[t_{\beta} > 1.645\right] = 0.058$$
$$\mathbb{P}\left[t_{\theta} > 1.645\right] = 0.006$$

Since $t_{\beta}, t_{\theta} \xrightarrow{\text{d}} \mathcal{N}(0,1)$, both probabilities should be close to 0.05. Thanks to the small sample size, it is not surprising that t_{θ} performs quite worse.

Why sample size matters? We further run simulations by varying sample size n: $50, 100, 500, 1000, 3000, 5000$, and investigate how the biases and probabilities evolve with those sample sizes.

From Fig. 1, both biases converge to 0, as sample sizes increase. In Fig. 2, the probability for t_{β} and t_{θ} hovers around 0.05, as the sample sizes become large. □

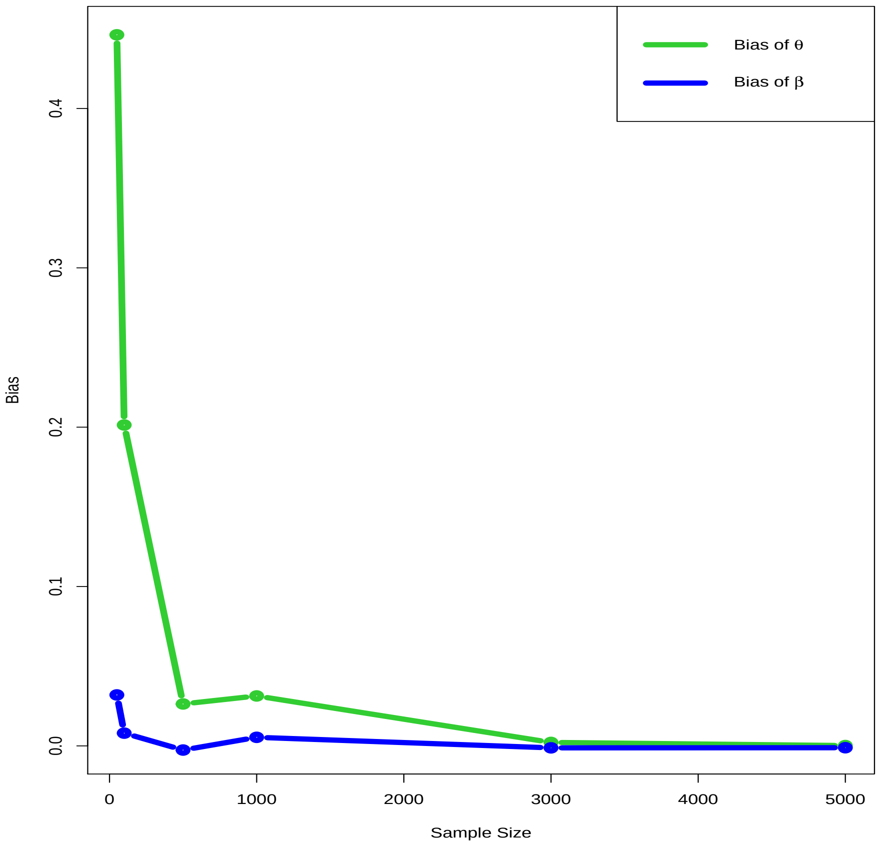

Figure 1: Bias $\left(\widehat{\beta}\right)$ and Bias $\left(\widehat{\theta}\right)$

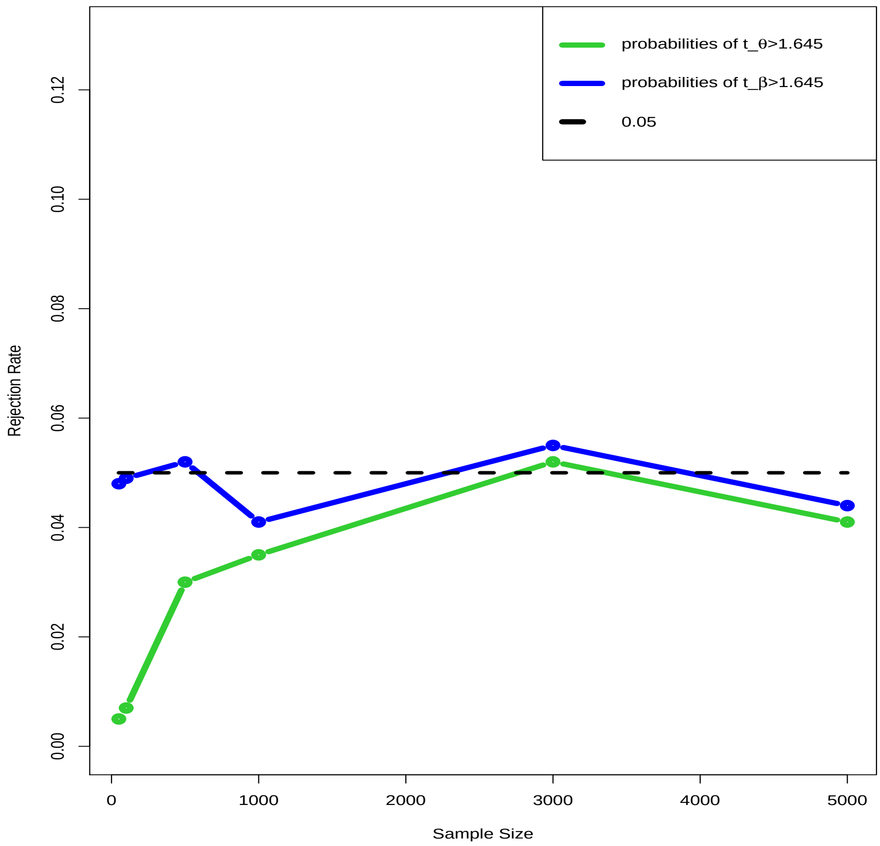

Figure 2: $\mathbb{P}\left[t_\beta > 1.645\right]$ and $\mathbb{P}\left[t_\theta > 1.645\right]$

R codes: 8.19

```
data <- read.csv("cps09mar.csv" ,header = TRUE,sep=",")
log_inc <- log(data$earnings/(data$hours*data$week))
edu <- data$education
exp_1 <- data$age - data$education- 6
exp_2 <- exp_1^2/100
married_1 <- ifelse(data$marital ==1, 1, 0)
married_2 <- ifelse(data$marital ==2, 1, 0)
married_3 <- ifelse(data$marital ==3, 1, 0)
widow <- ifelse(data$marital ==4 , 1, 0)
div <- ifelse(data$marital == 5, 1, 0)
sep <- ifelse(data$marital ==6, 1, 0)
subsample <- data$race ==1 & data$female ==0 & data$hisp == 1

## (a)
library(stargazer)
library(matlib)

Y = cbind(edu,exp_1,exp_2,married_1,married_2,married_3,widow,div,sep)
X = subset(Y,subsample)
y = log_inc[subsample]
n = dim(X)[1]
ones = rep(1,n)
k = 10

X_new = cbind(X,ones)
beta_ols = inv(t(X_new)%*%X_new)%*%t(X_new)%*%y
e_ols = y - X_new%*%beta_ols
Omega = matrix(0,nrow=k,ncol=k)
for (i in 1:n){
  Omega = X_new[i,]%*%t(X_new[i,])*(e_ols[i]^2) + Omega
 }
v_ols = inv(t(X_new)%*%X_new)%*%Omega%*%inv(t(X_new)%*%X_new)*n/(n-k)
se_ols = sqrt(diag(v_ols))

ols <- cbind(beta_ols,se_ols)
colnames(ols) = c("Estimates","se")
rownames(ols) = c("education","experience","experience 2","married 1","married 2",
    "married 3","widow","divorced","sep","intercept")
stargazer(ols)
## b

R = cbind(c(0,0,0,1,0,0,-1,0,0,0),c(0,0,0,0,0,0,0,1,-1,0))
```

```
beta_cls = beta_ols - inv(t(X_new)%*%X_new)%*%R%*%inv(t(R)%*%inv(t(X_new)%*%X_new)
    %*%R)%*%t(R)%*%beta_ols
e = y - X_new%*%beta_cls
e2 = e^2
s_cls2 = sum(e2)/(n-k-2)
X_inv = inv(t(X_new)%*%X_new)
R_inv = inv(t(R)%*%X_inv%*%R)
var_cls = (X_inv - X_inv%*%R%*%R_inv%*%t(R)%*%X_inv)*s_cls2
se_cls = sqrt(diag(var_cls))

cls <- cbind(beta_cls,se_cls)
colnames(cls) = c("Estimates","se")
rownames(cls) = c("education","experience","experience 2","married 1","married 2",
    "married 3","widow","divorced","sep","intercept")
stargazer(cls)
## b

## c

beta_emd = beta_ols - v_ols%*%R%*%inv(t(R)%*%v_ols%*%R)%*%t(R)%*%beta_ols
v_emd = v_ols - v_ols%*%R%*%inv(t(R)%*%v_ols%*%R)%*%t(R)%*%v_ols
se_emd = sqrt(diag(v_emd))

emd <- cbind(beta_emd,se_emd)
colnames(emd) = c("Estimates","se")
rownames(emd) = c("education","experience","experience 2","married 1","married 2",
    "married 3","widow","divorced","sep","intercept")
stargazer(emd)

## d
## e

R2 = cbind(R,c(0,1,1,0,0,0,0,0,0,0))

beta_inequal = beta_ols - v_ols%*%R2%*%inv(t(R2)%*%v_ols%*%R2)%*%t(R2)%*%beta_ols
v_inequal = v_ols - v_ols%*%R2%*%inv(t(R2)%*%v_ols%*%R2)%*%t(R2)%*%v_ols
se_inequal = sqrt(diag(v_inequal))

inequality <- cbind(beta_inequal,se_inequal)
colnames(inequality) = c("Estimates","se")
rownames(inequality) = c("education","experience","experience 2","married 1","
    married 2","married 3","widow","divorced","sep","intercept")
stargazer(inequality)
```

R codes: 9.24

```
library(sandwich)
n = 50

beta = 1
theta = exp(beta)

B = 1000

beta_est = rep(NA,B)
theta_est = rep(NA,B)
t_beta = rep(NA,B)
t_theta = rep(NA,B)
se_beta = rep(NA,B)
ones = rep(1,n)

for (i in 1:B){
 x = runif(n, min = 0, max = 1)
 e = rnorm(n,mean = 0, sd = 1)
 alpha = runif(1,-100,100)
 y = alpha*ones + beta*x + e
 fit <- lm(y~x)
 beta_est[i] = fit$coefficients[2]
 se_beta[i] = sqrt(vcovHC(fit,type="HC1")[2,2])
 theta_est[i] = exp(beta_est[i])
 t_beta[i] = (beta_est[i] - beta)/se_beta[i]
 se_theta = se_beta[i]*exp(beta_est[i])
   t_theta[i] = (theta_est[i] - theta)/se_theta
}
bias_beta = mean(beta_est) -beta
bias_theta = mean(theta_est) -theta
bias_beta
bias_theta
sum(t_beta > 1.645)/B
sum(t_theta > 1.645)/B


## Asymptotic Analysis

library(sandwich)
beta = 1
theta = exp(beta)

```

```r

B = 1000
beta_est = rep(NA,B)
theta_est = rep(NA,B)
t_beta = rep(NA,B)
t_theta = rep(NA,B)
sample_seq = c(50,100,500,1000,3000,5000)

Bias_beta = rep(NA,length(sample_seq))
Bias_theta = rep(NA,length(sample_seq))
Prob_beta = rep(NA,length(sample_seq))
Prob_theta = rep(NA,length(sample_seq))

for (i in 1:length(sample_seq)){
  n = sample_seq[i]
  for (j in 1:B){
    x = runif(n, min = 0, max = 1)
    e = rnorm(n,mean = 0, sd = 1)
    ones = rep(1,n)
    alpha = runif(1,-100,100)
    y = alpha*ones + beta*x + e
    fit <- lm(y~x)
    beta_est[j]= fit$coefficients[2]

    se_beta = sqrt(vcovHC(fit,type="HC1")[2,2])
    t_beta[j] = (beta_est[j] - beta)/se_beta
    theta_est[j] = exp(beta_est[j])
    se_theta = se_beta*exp(beta_est[j])
    t_theta[j] = (theta_est[j] - theta)/se_theta
  }
  Bias_beta[i] = mean(beta_est) - beta
  Bias_theta[i] = mean(theta_est) - theta
  Prob_beta[i] = sum(t_beta > 1.645)/B
  Prob_theta[i] = sum(t_theta>1.645)/B
}
plot(sample_seq,Bias_theta, type="b",lwd=6,col="limegreen",xlab="Sample Size",ylab
    ="Bias")

lines(sample_seq,Bias_beta, type="b",lwd=4,col="blue")

legend("topright", c(expression(paste("Bias of ", theta,)), expression(paste("Bias
     of ", beta)) ), lwd=6, col=c("limegreen","blue"))

strght_line = rep(0.05,length(sample_seq))
```

```
plot(sample_seq,Prob_theta, type="b",lwd=6,col="limegreen",xlab="Sample Size",ylab
    ="Rejection Rate",ylim=c(0.03,0.13))

lines(sample_seq,Prob_beta, type="b",lwd=6,col="blue")

lines(sample_seq,strght_line,lwd=4,lty=2,type="l")

legend("topright", c(expression(paste("probabilities of ", "t_",,theta,">1.645")),
     expression(paste("probabilities of ", "t_",beta,">1.645")) ,"0.05"), lwd=6,
    lty=c(1,1,2) ,col=c("limegreen","blue","black"))
```