# MEI MT Problem Set 6[1]

Timothee Carayol

November 26, 2009

[1] Available on http://personal.lse.ac.uk/carayolt/ec402.htm

# Question 1

## Question

- Set of four dummy variables $d1, d2, d3, d4$. $dk$ is $T \times 1$.
- $$\forall i \in \{1, .., T\}, k \in \{1, 2, 3, 4\}, \begin{cases} dk_i = 0 \; or \; 1 \\ \sum_k dk_i = 1 \end{cases}$$
- i.e, one and only one of the $dk$'s takes value 1 for each observation, the others are 0.

## Answer

- Suppose true model is
  $h_i = \beta_0 + \beta_1 d1_i + \beta_2 d2_i + \beta_3 d3_i + \beta_4 d4_i + x_i' \gamma + \epsilon_i$

# Question 1 (a)

## Answer

- Then the matrix of regressors is

$$X = (i \; d1 \; d2 \; d3 \; d4 \; X_2)$$

- Problem? $i = d1 + d2 + d3 + d4 \Rightarrow$ the columns of $X$ are not linearly independent.
- Hence $rank(X) < k$: $A1$ is violated, and we cannot compute $\hat{\beta}_{ols}$. ($X'X$ not invertible.)
- This is known as the "dummy variable trap". Solution: drop either the constant or one dummy.

# Question 1 (b)

## Question

- Want to compare two specifications $(i)$ and $(ii)$:
- $(i)$ $h_i = \beta_0 + \beta_2 d2_i + \beta_3 d3_i + \beta_4 d4_i + x_i'\gamma + \epsilon_i = x_i'\beta + \epsilon_i$
- $(ii)$ $h_i = \theta_1 d1_i + \theta_2 d2_i + \theta_3 d3_i + \theta_4 d4_i + x_i'\delta + \epsilon_i = z_i'\theta + \epsilon_i$

## Answer

- With: $X = (i\ d2\ d3\ d4\ X_2)$, $Z = (d1\ d2\ d3\ d4\ X_2)$, $\beta = \begin{pmatrix} \beta_0 & \beta_2 & \beta_3 & \beta_4 & \gamma \end{pmatrix}'$, $\theta = \begin{pmatrix} \theta_1 & \theta_2 & \theta_3 & \theta_4 & \delta \end{pmatrix}'$.
- Note that $Z = XA$ with
$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & 0_{k-4}' \\ -1 & 1 & 0 & 0 & 0_{k-4}' \\ -1 & 0 & 1 & 0 & 0_{k-4}' \\ -1 & 0 & 0 & 1 & 0_{k-4}' \\ 0_{k-4} & 0_{k-4} & 0_{k-4} & 0_{k-4} & I_{k-4} \end{pmatrix},\ k \times k$$
non-singular.

# Question 1 (b) (cont)

## Answer

- So from last week, we know that computing OLS on either specification will yield the same predicted values and residuals.
- We also know from last week that $\hat{\theta}_{ols} = A^{-1}\hat{\beta}_{ols}$, i.e. $\hat{\beta}_{ols} = A\hat{\theta}_{ols}$.

$$\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \\ \hat{\gamma} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0'_{k-4} \\ -1 & 1 & 0 & 0 & 0'_{k-4} \\ -1 & 0 & 1 & 0 & 0'_{k-4} \\ -1 & 0 & 0 & 1 & 0'_{k-4} \\ 0_{k-4} & 0_{k-4} & 0_{k-4} & 0_{k-4} & I_{k-4} \end{pmatrix} \begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \\ \hat{\theta}_3 \\ \hat{\theta}_4 \\ \hat{\delta} \end{pmatrix}$$

$$= \begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 - \hat{\theta}_1 \\ \hat{\theta}_3 - \hat{\theta}_1 \\ \hat{\theta}_4 - \hat{\theta}_1 \\ \hat{\delta} \end{pmatrix}$$

# Question 2

## Question

- Similar setup as previously. Set of four dummy variables $d1, d2, d3, d4$. $dk$ is $T \times 1$.
- True model for $y$: $y = S\beta_1 + X_2\beta_2 + \epsilon$ with $S = (d1\ d2\ d3\ d4)$, where each $dk$ is a quarter dummy.
- Purpose of the exercise: show that the OLS estimator from the specification above is numerically equal to the OLS estimator obtained from regressing the "deseasonalized" $y$ on the "deseasonalized" $X_2$.
- "Deseasonalize" means removing the seasonal variations from a variable, i.e. take the residuals from the regression of that variable on the quarter dummies.
- (a) Show that for any $T \times 1$ vector $z$, the formula for the "deseasonalized" $z$ is therefore $M_S z$.

# Question 2 (a)

## Answer

- $M_S z = (I_T - S(S'S)^{-1}S')z$.
- $S'S = \begin{pmatrix} d1'd1 & d1d'2 & d1'd3 & d1'd4 \\ d2'd1 & d2'd2 & d2'd3 & d2'd4 \\ d3'd1 & d3'd2 & d3'd3 & d3'd4 \\ d4'd1 & d4'd2 & d4'd3 & d4'd4 \end{pmatrix}$
- $\forall i \neq j,\ (i,j) \in 1,2,3,4^2,\ di'dj = \sum_{t=1}^{T} di_t dj_t = 0$ because either $di_t$ or $dj_t$ (or both) are zero.
- $\forall i \in 1,2,3,4^2,\ di'di = \sum_{t=1}^{T} di_t^2 = \sum_{t=1}^{T} di_t$ (because $di_t$ is 0 or 1, hence equals $di_t^2$).
- Let us denote $T_i = \sum_{t=1}^{T} di_t$. $T_i$ is the number of observations in quarter $i$, i.e. such that $di_t = 1$.

# Question 2 (a) (cont)

## Answer

- $S'S = \begin{pmatrix} T_1 & 0 & 0 & 0 \\ 0 & T_2 & 0 & 0 \\ 0 & 0 & T_3 & 0 \\ 0 & 0 & 0 & T_4 \end{pmatrix} \Rightarrow (S'S)^{-1} = \begin{pmatrix} \frac{1}{T_1} & 0 & 0 & 0 \\ 0 & \frac{1}{T_2} & 0 & 0 \\ 0 & 0 & \frac{1}{T_3} & 0 \\ 0 & 0 & 0 & \frac{1}{T_4} \end{pmatrix}$
- $S'z = \begin{pmatrix} d1'z \\ d2'z \\ d3'z \\ d4'z \end{pmatrix} = \begin{pmatrix} \sum_{t=1}^{T} d1_t z_t \\ \sum_{t=1}^{T} d2_t z_t \\ \sum_{t=1}^{T} d3_t z_t \\ \sum_{t=1}^{T} d4_t z_t \end{pmatrix}$

# Question 2 (a) (cont)

## Answer

- Hence:

$$(S'S)^{-1}S'z = \begin{pmatrix} \frac{1}{T_1} & 0 & 0 & 0 \\ 0 & \frac{1}{T_2} & 0 & 0 \\ 0 & 0 & \frac{1}{T_3} & 0 \\ 0 & 0 & 0 & \frac{1}{T_4} \end{pmatrix} \begin{pmatrix} \sum_{t=1}^{T} d1_t z_t \\ \sum_{t=1}^{T} d2_t z_t \\ \sum_{t=1}^{T} d3_t z_t \\ \sum_{t=1}^{T} d4_t z_t \end{pmatrix} \quad (1)$$

$$= \begin{pmatrix} \frac{\sum_{t=1}^{T} d1_t z_t}{T_1} \\ \frac{\sum_{t=1}^{T} d2_t z_t}{T_2} \\ \frac{\sum_{t=1}^{T} d3_t z_t}{T_3} \\ \frac{\sum_{t=1}^{T} d4_t z_t}{T_4} \end{pmatrix} = \begin{pmatrix} \overline{z}^{(1)} \\ \overline{z}^{(2)} \\ \overline{z}^{(3)} \\ \overline{z}^{(4)} \end{pmatrix} \quad (2)$$

where $\overline{z}^{(k)}$ is the mean of variable $z$ for all observations in quarter $k$.

# Question 2 (a) (cont)

## Answer

- Hence the $t$th row of $S(S'S)^{-1}S'z$ is $\overline{z}^{(q_t)}$ with $q_t = i$ iff $di_t = 1$ (e.g. $\overline{z}^{(1)}$ for an observation from the first quarter).
- Thus, the $t$th row of $M_S z$ is $(I_T - S(S'S)^{-1}S')z_t = z_t - \overline{z}^{(q_t)}$, that is to say, $z_t$ minus the seasonal average for $t$th quarter. (QED)

# Question 2 (b)

## Question

Explain why $\hat{\beta}_{2OLS}$ can also be obtained from regressing deseasonalized $y$ on deseasonalized $X_2$.

## Answer

- Note that, from Frisch-Waugh theorem, $\hat{\beta}_2 = (X_2' M_S X_2)^{-1} X_2' M_S y = ((M_S X_2)' M_S X_2)^{-1} (M_S X_2)^{-1} M_S y$.
- Hence, $\hat{\beta}_2$ is the OLS estimator from the regression of deseasonalized $y$ on deseasonalized $X_2$.

# Question 3 (a) i.

## Question

- Exact same setup as question 1 from PS5, so we will use some results from that question directly here. I assume A3>A3Rsru.
- (a) i. We want to test $\begin{cases} H_0 : \beta_1 = 0 \\ H_1 : \beta_1 \neq 0 \end{cases}$

## Answer

- We know that under A1-A5,
$\hat{\beta}|X \sim \mathcal{N}(\beta, \sigma_\epsilon^2 (X'X)^{-1}) = \mathcal{N}\left(\beta, \frac{\sigma_\epsilon^2}{3} \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}\right)$.
- So $Z_1 = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{2}{3}\sigma_\epsilon^2}} \sim \mathcal{N}(0, 1)$

# Question 3 (a) i. (cont)

## Answer

- We can therefore show that

$$T_1 = \frac{\hat{\beta}_1 - \beta_1}{SE(\hat{\beta}_1)} = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{2}{3}s^2}} = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{4}{81}}} = \frac{\frac{1}{3} - \beta_1}{\frac{2}{9}} \sim t(N-K) = t(9)$$

- In this question: want to test $\begin{cases} H_0 : \beta_1 = 0 \\ H_1 : \beta_1 \neq 0 \end{cases}$ Under $H_0$,

  $T_1 = \frac{\frac{1}{3}}{\frac{2}{9}} = 1.5 < t_{2.5\%}(9) = 2.262$, so we do not reject $H_0$.

# Question 3 (a) ii.

## Answer

- Very similar.
- $Z_2 = \frac{\hat{\beta}_2 - \beta_2}{\sqrt{\frac{2}{3}\sigma_\epsilon^2}} \sim \mathcal{N}(0,1)$: $\beta_2$, here, has the same standard deviation as $\beta_1$.
- So
$$T_2 = \frac{\frac{1}{3} - \beta_2}{\frac{2}{9}} \sim t(N-K) = t(9)$$
- In this question: want to test $\begin{cases} H_0 : \beta_2 = 0 \\ H_1 : \beta_2 \neq 0 \end{cases}$ Under $H_0$, $T_2 = \frac{\frac{1}{3}}{\frac{2}{9}} = 1.5 < t_{2.5\%}(9) = 2.262$, so we do not reject $H_0$.

# Question 3 (a) iii.

## Question

- (a) We want to test $\begin{cases} H_0 : \beta = \begin{pmatrix} 0 & 0 \end{pmatrix}' \\ H_1 : \beta \neq \begin{pmatrix} 0 & 0 \end{pmatrix}' \end{cases}$

## Answer

- Perform Wald test: use the fact that, if $R$ is $r \times K$, $F = \frac{(R\hat{\beta}-R\beta)'\left(s^2R(X'X)^{-1}R'\right)^{-1}(R\hat{\beta}-R\beta)}{r} \sim F(r, N-K)$.
- Here: $R = I_2$; under $H_0$, $R\beta = 0_2$. Therefore under $H_0$, $F = \frac{\hat{\beta}'\left(s^2(X'X)^{-1}\right)^{-1}\hat{\beta}}{2} \sim F(2, 9)$.
- So here, $F = \frac{27}{4}(\frac{1}{3}\,\frac{1}{3})\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} \frac{1}{3} \\ \frac{1}{3} \end{pmatrix} = 4.5.$
- Compare this to $F_{5\%}(2,9) = 4.26 < F$: hence we can reject $H_0$ at the 5% level. Our estimates are jointly significant (even thought they are not individually significant).

# Question 3 (b)

## Answer

- Rather subtle point: sometimes testing joint significance gives different results than succesive tests of individual significances. This happens typically when the $X$'s are correlated (but not perfectly, otherwise OLS could not even be computed). e.g. imagine $x_1$ and $x_2$ strongly positively correlated. Coefficient on $x_1$ indicates what happens to the dependent variable when $x_1$ varies but $x_2$ remains constant. If $x_1$ and $x_2$ typically vary together, then there is much less information we can use than if they are independent: hard to tell how much of the variations in $y$ are due to variations in $x_1$ and how much to variations in $x_2$. In extreme case of perfect collinearity, we cannot even compute $\hat{\beta}_{ols}$.

# Question 4

## Question

$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + \epsilon_t, t = 1, .., T$ with $A1, A2, A3Rmi, A4GM, A5normal$. (a) Rewrite 3 sets of hypotheses in the form $R\beta = q$, with $R$ $r \times 4$. (b) Rewrite the $W$ statistic in the form $F = \frac{\frac{RSS_R - RSS_U}{r}}{\frac{RSS_U}{T-K}}$. Explain how you would find $RSS_R$.

## Answer

- $(i) H_0 : \beta_2 - 3\beta_3 = 4, \beta_1 = 2\beta_4 \Rightarrow r = 2, q = \begin{pmatrix} 4 \\ 0 \end{pmatrix}, R =$
  $\begin{pmatrix} 0 & 1 & -3 & 0 \\ 1 & 0 & 0 & -2 \end{pmatrix}$, and
  $W = (R\hat{\beta} - q)' \left( R\hat{V}(\hat{\beta})R' \right)^{-1} (R\hat{\beta} - q)/r \sim F(2, T-4).$

# Question 4 (a)

## Answer

- $(ii) H_0 : \beta_1 = 1, \beta_2 = 3\beta_4 - 1, \beta_3 = 0 \Rightarrow r = 3, q = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, R = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -3 \\ 0 & 0 & 1 & 0 \end{pmatrix}$, and $W \sim F(3, T-4)$.
- $(iii) H_0 : \beta_2 = \beta_3 = 2\beta_4 \Rightarrow r = 2, q = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, R = \begin{pmatrix} 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -2 \end{pmatrix}$, and $W \sim F(2, T-4)$.

# Question 4 (b)

## Answer

- (i) Obtain $RSS_R$ from $y_t = 2\beta_4 + (3\beta_3 + 4)x_{2t} + \beta_3 x_{3t} \beta_4 x_{4t} + \epsilon_t$, or $y_t - 4x_{2t} = \beta_3(3x_{2t} + x_{3t}) + \beta_4(2 + x_{4t}) + \epsilon_t$—i.e., regress $y_t - 4x_{2t}$ on $3x_{2t} + x_{3t}$ and $2 + x_{4t}$ without a constant (because we constrained the constant $\beta_1$ to be equal to twice the coefficient on $x_4$).
- (ii) Same idea $\Rightarrow$ Regress $y_t - 1 + x_{2t}$ on $3x_{2t} + x_{4t}$ without a constant (we constrained the constant to be equal to 1).
- (iii) Same idea $\Rightarrow$ Regress $y_t$ on a constant (because we haven't put any constraint on $\beta_1$ this time) and $2x_{2t} + 2x_{3t} + x_{4t}$.