Lecture 5-6: Serial correlation

Econometric Methods

Andrzej Torój

SGH Warsaw School of Economics – Institute of Econometrics

Outline

Outline

Variance-covariance matrix of $\boldsymbol{\varepsilon}$

$$E\left(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T\right) = E\left(\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{bmatrix}\begin{bmatrix} \varepsilon_1 & \varepsilon_2 & \dots & \varepsilon_T \end{bmatrix}\right) = E\left(\begin{bmatrix} \varepsilon_1^2 & \varepsilon_1\varepsilon_2 & \dots & \varepsilon_1\varepsilon_T \\ \varepsilon_1\varepsilon_2 & \varepsilon_2^2 & \dots & \varepsilon_2\varepsilon_T \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon_1\varepsilon_T & \varepsilon_2\varepsilon_T & \dots & \varepsilon_T^2 \end{bmatrix}\right) =$$

$$= \begin{bmatrix} E\left(\varepsilon_1 - E\left(\varepsilon_1\right)\right)^2 & E\left(\varepsilon_1 - E\left(\varepsilon_1\right)\right)\left(\varepsilon_2 - E\left(\varepsilon_2\right)\right) & \dots & E\left(\varepsilon_1 - E\left(\varepsilon_1\right)\right)\left(\varepsilon_T - E\left(\varepsilon_T\right)\right) \\ E\left(\varepsilon_1 - E\left(\varepsilon_1\right)\right)\left(\varepsilon_2 - E\left(\varepsilon_2\right)\right) & E\left(\varepsilon_2 - E\left(\varepsilon_2\right)\right)^2 & \dots & E\left(\varepsilon_2 - E\left(\varepsilon_2\right)\right)\left(\varepsilon_T - E\left(\varepsilon_T\right)\right) \\ \vdots & \vdots & \ddots & \vdots \\ E\left(\varepsilon_1 - E\left(\varepsilon_1\right)\right)\left(\varepsilon_T - E\left(\varepsilon_T\right)\right) & E\left(\varepsilon_2 - E\left(\varepsilon_2\right)\right)\left(\varepsilon_T - E\left(\varepsilon_T\right)\right) & \dots & E\left(\varepsilon_T - E\left(\varepsilon_T\right)\right)^2 \end{bmatrix}$$

$$= \begin{bmatrix} var\left(\varepsilon_1\right) & cov\left(\varepsilon_1\varepsilon_2\right) & \dots & cov\left(\varepsilon_1\varepsilon_T\right) \\ cov\left(\varepsilon_1\varepsilon_2\right) & var\left(\varepsilon_2\right) & \dots & cov\left(\varepsilon_2\varepsilon_T\right) \\ \vdots & \vdots & \ddots & \vdots \\ cov\left(\varepsilon_1\varepsilon_T\right) & cov\left(\varepsilon_2\varepsilon_T\right) & \dots & var\left(\varepsilon_T\right) \end{bmatrix} \overset{\text{OLS assumptions}}{=} \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix}$$

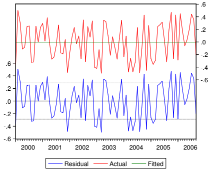
Spherical disturbances

$$\begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix} \text{ – a spherical matrix}$$

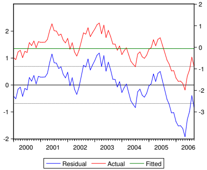
- more general than identity matrix
- more restrictive than diagonal matrix
  - zero non-diagonal elements (if broken: **serial correlation**)
  - constant diagonal elements (if broken: **heteroskedasticity**)

Serial correlation in data

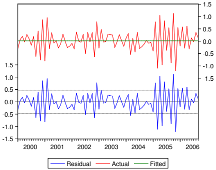
no serial correlation



positive



negative



Does serial correlation cause *bias*?

$$\begin{aligned} E\left(\hat{\beta}\right) &= E\left(\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}\right) = \\ &= E\left(\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\left(\mathbf{X}\beta+\varepsilon\right)\right) = \\ &= E\left(\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{X}\beta\right) + E\left(\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\varepsilon\right) = \\ &= \beta + \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T E\left(\varepsilon\right) = \beta \end{aligned}$$

- unbiasedness of OLS proved – did we use the assumption of spherical disturbances ε?

- NO!
- ...but note that
  - serial correlation can be a symptom of misspecification...
  - ...which in turn can bias the estimates

- NO!
- ...but note that
  - serial correlation can be a symptom of misspecification...
  - ...which in turn can bias the estimates

Does serial correlation cause *inefficiency*?

Recall: $\hat{\beta} = \beta + \left(\mathbf{X}^T\mathbf{X}\right)\mathbf{X}^T\varepsilon$ (do the previous derivation without the expected value operator)

$$\begin{aligned} Var\left(\hat{\beta}\right) &= E\left[\left(\hat{\beta}-\beta\right)\left(\hat{\beta}-\beta\right)^T\right] = \\ &= E\left(\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\varepsilon\cdot\varepsilon^T\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\right) = \\ &= \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T E\left(\varepsilon\varepsilon^T\right)\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1} = \dots \end{aligned}$$

- under spherical disturbances: $E\left(\varepsilon\varepsilon^T\right) = \sigma^2\mathbf{I}$
  $\dots = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\sigma^2\mathbf{I}\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1} = \sigma^2\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1} = \sigma^2\left(\mathbf{X}^T\mathbf{X}\right)^{-1}$
  (square roots of diagonal elements are standard errors of estimation)
- under non-spherical disturbances: $E\left(\varepsilon\varepsilon^T\right) = \mathbf{\Omega}$
  $\dots = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{\Omega}\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}$

Does serial correlation cause *inefficiency*?

Recall: $\hat{\beta} = \beta + (\mathbf{X}^T\mathbf{X})\mathbf{X}^T\varepsilon$ (do the previous derivation without the expected value operator)

$$\begin{aligned} Var\left(\hat{\beta}\right) &= E\left[\left(\hat{\beta}-\beta\right)\left(\hat{\beta}-\beta\right)^T\right] = \\ &= E\left(\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\varepsilon\cdot\varepsilon^T\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\right) = \\ &= \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T E\left(\varepsilon\varepsilon^T\right)\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1} = \ldots \end{aligned}$$

- under spherical disturbances: $E\left(\varepsilon\varepsilon^T\right) = \sigma^2\mathbf{I}$
  $\ldots = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\sigma^2\mathbf{I}\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1} = \sigma^2\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1} = \sigma^2\left(\mathbf{X}^T\mathbf{X}\right)^{-1}$
  (square roots of diagonal elements are standard errors of estimation)
- under non-spherical disturbances: $E\left(\varepsilon\varepsilon^T\right) = \mathbf{\Omega}$
  $\ldots = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{\Omega}\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}$

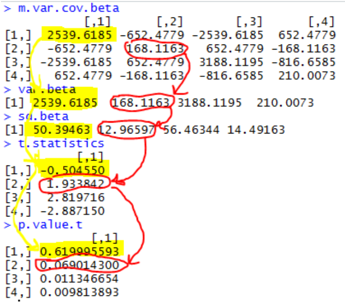
Does serial correlation cause *inefficiency*?

Recall: $\hat{\beta} = \beta + \left(\mathbf{X}^T\mathbf{X}\right)\mathbf{X}^T\varepsilon$ (do the previous derivation without the expected value operator)

$$
\begin{aligned}
Var\left(\hat{\beta}\right) &= E\left[\left(\hat{\beta}-\beta\right)\left(\hat{\beta}-\beta\right)^T\right] = \\
&= E\left(\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\varepsilon \cdot \varepsilon^T\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\right) = \\
&= \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T E\left(\varepsilon\varepsilon^T\right)\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1} = \ldots
\end{aligned}
$$

- under spherical disturbances: $E\left(\varepsilon\varepsilon^T\right) = \sigma^2\mathbf{I}$
  $\ldots = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\sigma^2\mathbf{I}\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1} = \sigma^2\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1} = \sigma^2\left(\mathbf{X}^T\mathbf{X}\right)^{-1}$
  (square roots of diagonal elements are standard errors of estimation)
- under non-spherical disturbances: $E\left(\varepsilon\varepsilon^T\right) = \mathbf{\Omega}$
  $\ldots = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{\Omega}\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}$

- YES! The assumption of spherical matrix used explicitly in the derivation!
- If broken...
  - loss of efficiency
  - distorted statistical inference based on variance-covariance matrix of the error term, including *t*-tests of variable significance – recall from previous lecture:



```
> m.var.cov.beta
            [,1]       [,2]       [,3]      [,4]
[1,]  2539.6185  -652.4779 -2539.6185  652.4779
[2,]  -652.4779   168.1163   652.4779 -168.1163
[3,] -2539.6185   652.4779  3188.1195 -816.6585
[4,]   652.4779  -168.1163  -816.6585  210.0073
> var.beta
[1] 2539.6185  168.1163 3188.1195  210.0073
> sd.beta
[1] 50.39463 12.96597 56.46344 14.49163
> t.statistics
          [,1]
[1,] -0.504550
[2,]  1.933842
[3,]  2.819716
[4,] -2.887150
> p.value.t
            [,1]
[1,] 0.619995593
[2,] 0.069014300
[3,] 0.011346654
[4,] 0.009813893
```

Sources of serial correlation

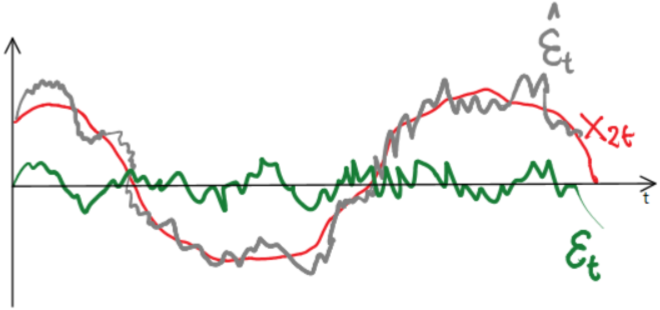
- inertia and persistence of economic phenomena:
  - **serial correlation approach** (accept the serial correlation and try to improve the efficiency and statistical inference)
- specification error (misspecified functional form, misspecified dynamics, omitted variables)
  - **re-specification approach** (modify the model specification until serial correlation vanishes)

Sources of serial correlation

- inertia and persistence of economic phenomena:
  - **serial correlation approach** (accept the serial correlation and try to improve the efficiency and statistical inference)
- specification error (misspecified functional form, misspecified dynamics, omitted variables)
  - **re-specification approach** (modify the model specification until serial correlation vanishes)

Re-specification approach: idea

"True" model:	$y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \varepsilon_t$
Estimated model:	$y_t = \hat{\beta}_0 + \hat{\beta}_1 x_{1,t} + \hat{\varepsilon}_t$



Exercise (1/3)

Gasoline demand model

Our regression model explains the log-demand for gasoline per capita (***log_Q_gasoline***) with the following variables: log fuel price (***log_P_gasoline***), log per capita income (***log_Income***) and log prices of new and used cars (***log_P_new_car*** and ***log_P_used_car***, respectively). Model should contain a constant.

1. What is the consequence of using logged variables for the interpretation of the regression results?
2. Which signs of the estimates do You expect? How do these expectations relate to familiar microeconomic theories?
3. Write down the estimated equation.
4. Were the expectations from point 2 fulfilled?

Outline

Durbin-Watson (1)

Durbin-Watson test

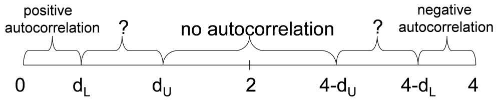
H_0 : error term has no serial correlation of order 1

H_1 : error term does have serial correlation of order 1

$$DW = \frac{\sum_{t=2}^{T} (\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum_{t=1}^{T} \hat{\varepsilon}_t^2} \approx 2(1-\rho)$$

where ρ – correlation between $\hat{\varepsilon}_t$ and $\hat{\varepsilon}_{t-1}$

Test statistics has its own distribution tables.



Durbin-Watson (2)

Limitations:

- model with a constant
- ...without lagged dependent variable
- ...with normally distributed error term
- detects serial correlation of order at most 1
- inconclusive range

Durbin-Watson (2)

Limitations:

- model with a constant
- **...without lagged dependent variable**
- ...with normally distributed error term
- **detects serial correlation of order at most 1**
- inconclusive range

Durbin-Watson (2)

Limitations:

- model with a constant
- **...without lagged dependent variable**
- ...with normally distributed error term
- **detects serial correlation of order at most 1**
- inconclusive range

Durbin-Watson (2)

Limitations:

- model with a constant
- **...without lagged dependent variable**
- ...with normally distributed error term
- **detects serial correlation of order at most 1**
- inconclusive range

Durbin-Watson (2)

Limitations:

- model with a constant
- **...without lagged dependent variable**
- ...with normally distributed error term
- **detects serial correlation of order at most 1**
- inconclusive range

h-Durbin

- DW test biased towards rejecting serial correlation when lagged dependent variable is a regressor
- see e.g. *Nerlove, Wallis (1966)*
- Durbin proposed a modification to take account of that:

h-Durbin test

H_0 : error term has no serial correlation of order 1

H_1 : error term does have serial correlation of order 1

$$hd = \left(1 - \frac{DW}{2}\right)\sqrt{\frac{T}{1 - T \cdot var\left(\hat{\beta}_{y(t-1)}\right)}} \sim N(0,1)$$

Lagrange multiplier

- Step 1: **regression equation estimated via OLS**
  $y_t = \mathbf{x}_t\boldsymbol{\beta} + \varepsilon_t \qquad \hat{\boldsymbol{\beta}} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y} \qquad \hat{\varepsilon}_t = y_t - \mathbf{x}_t\hat{\boldsymbol{\beta}}$
- Step 2: **auxiliary regression for residuals from step 1**
  $\hat{\varepsilon}_t = \mathbf{x}_t\boldsymbol{\beta} + \beta_{k+1}\hat{\varepsilon}_{t-1} + \beta_{k+2}\hat{\varepsilon}_{t-2} + \ldots + \beta_{k+P}\hat{\varepsilon}_{t-P}$
  Under no serial correlation up to order P, the R^2 of auxiliary regression should be low.

Lagrange multiplier test

H_0 : error term has no serial correlation of order up to P
H_1 : error term does have serial correlation of order from 1 to P
$LM = TR^2 \sim \chi^2(P)$
Attention! Asymptotic test (short samples may distort the inference).

Lagrange multiplier

- Step 1: **regression equation estimated via OLS**
  $y_t = \mathbf{x}_t\boldsymbol{\beta} + \varepsilon_t \qquad \hat{\boldsymbol{\beta}} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y} \qquad \hat{\varepsilon}_t = y_t - \mathbf{x}_t\hat{\boldsymbol{\beta}}$
- Step 2: **auxiliary regression for residuals from step 1**
  $\hat{\varepsilon}_t = \mathbf{x}_t\boldsymbol{\beta} + \beta_{k+1}\hat{\varepsilon}_{t-1} + \beta_{k+2}\hat{\varepsilon}_{t-2} + \ldots + \beta_{k+P}\hat{\varepsilon}_{t-P}$
  Under no serial correlation up to order P, the R^2 of auxiliary regression should be low.

Lagrange multiplier test

H_0 : error term has no serial correlation of order up to P
H_1 : error term does have serial correlation of order from 1 to P
$LM = TR^2 \sim \chi^2(P)$
Attention! Asymptotic test (short samples may distort the inference).

Lagrange multiplier

- Step 1: **regression equation estimated via OLS**
  $y_t = \mathbf{x}_t \boldsymbol{\beta} + \varepsilon_t \qquad \hat{\boldsymbol{\beta}} = \left(\mathbf{X}^T \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{y} \qquad \hat{\varepsilon}_t = y_t - \mathbf{x}_t \hat{\boldsymbol{\beta}}$
- Step 2: **auxiliary regression for residuals from step 1**
  $\hat{\varepsilon}_t = \mathbf{x}_t \boldsymbol{\beta} + \beta_{k+1} \hat{\varepsilon}_{t-1} + \beta_{k+2} \hat{\varepsilon}_{t-2} + \ldots + \beta_{k+P} \hat{\varepsilon}_{t-P}$
  Under no serial correlation up to order P, the R^2 of auxiliary regression should be low.

Lagrange multiplier test

H_0 : error term has no serial correlation of order up to P
H_1 : error term does have serial correlation of order from 1 to P
$LM = TR^2 \sim \chi^2(P)$
Attention! Asymptotic test (short samples may distort the inference).

Ljung-Box

Ljung-Box test

H_0 : error term has no serial correlation of order up to P
H_1 : error term does have serial correlation of order from 1 to P

$$Q = T(T+2)\sum_{j=1}^{P}\left(\frac{1}{T-j}\cdot\frac{\sum_{t=j+1}^{T}\hat{\varepsilon}_t\hat{\varepsilon}_{t-j}}{\sum_{t=1}^{T}\hat{\varepsilon}_t^2}\right)$$

High values of Q imply serial correlation (reject the null).

Exercise (2/3)

Gasoline demand model

Look at our model.

1. Are the residuals serially correlated of order 1? Can we use Durbin-Watson test?
2. Are they serially correlated of order up to 2? Use LM and Ljung-Box test.
3. Is there serial correlation of order 2? Can we give a conclusive answer?
4. Would the residuals be serially correlated if we considered a model with additional lag of the dependent variable?

Outline

What are robust standard errors (and what they are not)?

- formula for variance-covariance matrix of $\hat{\beta}$, $Var\left(\hat{\beta}\right) = \hat{\sigma}^2 \left(\mathbf{X}^T\mathbf{X}\right)^{-1}$, **true only when errors are spherical**
- under serial correlation it is **false**, in particular:
  - the standard errors of estimation $s\left(\hat{\beta}_0\right)$ from the diagonal $Var\left(\hat{\beta}\right)$...
  - ...t-statistics $t = \frac{\hat{\beta}_i}{s\left(\hat{\beta}_1\right)}$...
  - ...and conclusions concerning the significance of variables
- robust standard errors – alternative way of estimating $s\left(\hat{\beta}_0\right)$

Choosing robust standard errors does not affect estimation results...

...but only the assessment of standard errors and the resulting conclusions as regards the significance of variables.

What are robust standard errors (and what they are not)?

- formula for variance-covariance matrix of $\hat{\boldsymbol{\beta}}$, $Var\left(\hat{\boldsymbol{\beta}}\right) = \hat{\sigma}^2 \left(\mathbf{X}^T\mathbf{X}\right)^{-1}$, **true only when errors are spherical**
- under serial correlation it is **false**, in particular:
  - the standard errors of estimation $s\left(\hat{\beta}_0\right)$ from the diagonal $Var\left(\hat{\boldsymbol{\beta}}\right)$...
  - ...t-statistics $t = \frac{\hat{\beta}_i}{s\left(\hat{\beta}_1\right)}$...
  - ...and conclusions concerning the significance of variables
- robust standard errors – alternative way of estimating $s\left(\hat{\beta}_0\right)$

Choosing robust standard errors does not affect estimation results...

...but only the assessment of standard errors and the resulting conclusions as regards the significance of variables.

What are robust standard errors (and what they are not)?

- formula for variance-covariance matrix of $\hat{\boldsymbol{\beta}}$, $Var\left(\hat{\boldsymbol{\beta}}\right) = \hat{\sigma}^2 \left(\mathbf{X}^T\mathbf{X}\right)^{-1}$, **true only when errors are spherical**
- under serial correlation it is **false**, in particular:
  - the standard errors of estimation $s\left(\hat{\beta}_0\right)$ from the diagonal $Var\left(\hat{\boldsymbol{\beta}}\right)$...
  - ...t-statistics $t = \frac{\hat{\beta}_i}{s\left(\hat{\beta}_1\right)}$...
  - ...and conclusions concerning the significance of variables
- robust standard errors – alternative way of estimating $s\left(\hat{\beta}_0\right)$

Choosing robust standard errors does not affect estimation results...

...but only the assessment of standard errors and the resulting conclusions as regards the significance of variables.

What are robust standard errors (and what they are not)?

- formula for variance-covariance matrix of $\hat{\beta}$, $Var\left(\hat{\beta}\right) = \hat{\sigma}^2 \left(\mathbf{X}^T\mathbf{X}\right)^{-1}$, **true only when errors are spherical**
- under serial correlation it is **false**, in particular:
  - the standard errors of estimation $s\left(\hat{\beta}_0\right)$ from the diagonal $Var\left(\hat{\beta}\right)$...
  - ...t-statistics $t = \frac{\hat{\beta}_i}{s\left(\hat{\beta}_1\right)}$...
  - ...and conclusions concerning the significance of variables
- robust standard errors – alternative way of estimating $s\left(\hat{\beta}_0\right)$

Choosing robust standard errors does not affect estimation results...

...but only the assessment of standard errors and the resulting conclusions as regards the significance of variables.

Newey-West robust SE

- ***Newey and West (1987)*** – robust to serial correlation (and heteroskedasticity):

$$Var\left(\hat{\beta}\right) = \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t^2 \mathbf{x}_t\mathbf{x}_t^T \left(X^T X\right)^{-1} + \\ + \frac{1}{T}\sum_{l=1}^{L}\sum_{t=l+1}^{T}\left(1-\frac{l}{L+1}\right)\hat{\varepsilon}_t\hat{\varepsilon}_{t-l}\left(\mathbf{x}_t\mathbf{x}_{t-l}^T + \mathbf{x}_{t-l}\mathbf{x}_t^T\right)$$

whereby L – arbitrary ($L \approx T^{\frac{1}{4}}$).

Generalised Least Squares estimator (1)

- $\mathbf{y} = \mathbf{X}\beta + \varepsilon \qquad \varepsilon \sim \left(E[\varepsilon] = \mathbf{0}, E\left[\varepsilon\varepsilon^T\right] = \sigma^2\mathbf{\Omega}\right)$

With well-defined, symmetric matrix $\mathbf{\Omega}$...

...there is a unique decomposition $\mathbf{\Omega}^{-1} = \mathbf{V}^T\mathbf{V}$.

- Pre-multiply both sides of equation with $\mathbf{V}$: $\mathbf{V}\mathbf{y} = \mathbf{V}\mathbf{X}\beta + \mathbf{V}\varepsilon$
- The error term in this model is spherical:

$$\begin{aligned} var(\mathbf{V}\varepsilon) &= E\left[(\mathbf{V}\varepsilon)(\mathbf{V}\varepsilon)^T\right] = E\left[\mathbf{V}\varepsilon\varepsilon^T\mathbf{V}^T\right] = \\ &= \mathbf{V}E\left[\varepsilon\varepsilon^T\right]\mathbf{V}^T = \mathbf{V}\sigma^2\mathbf{\Omega}\mathbf{V}^T = \\ &= \sigma^2\mathbf{V}\left(\mathbf{V}^T\mathbf{V}\right)^{-1}\mathbf{V}^T = \sigma^2\mathbf{V}\mathbf{V}^{-1}\left(\mathbf{V}^T\right)^{-1}\mathbf{V}^T = \sigma^2\mathbf{I} \end{aligned}$$

Generalised Least Squares estimator (1)

- $\mathbf{y} = \mathbf{X}\beta + \varepsilon \qquad \varepsilon \sim \left(E\left[\varepsilon\right] = \mathbf{0}, E\left[\varepsilon\varepsilon^T\right] = \sigma^2\mathbf{\Omega}\right)$

With well-defined, symmetric matrix $\mathbf{\Omega}$...

...there is a unique decomposition $\mathbf{\Omega}^{-1} = \mathbf{V}^T\mathbf{V}$.

- Pre-multiply both sides of equation with $\mathbf{V}$:
  $\mathbf{Vy} = \mathbf{VX}\beta + \mathbf{V}\varepsilon$
- The error term in this model is spherical:

$$\begin{aligned} var\left(\mathbf{V}\varepsilon\right) &= E\left[\left(\mathbf{V}\varepsilon\right)\left(\mathbf{V}\varepsilon\right)^T\right] = E\left[\mathbf{V}\varepsilon\varepsilon^T\mathbf{V}^T\right] = \\ &= \mathbf{V}E\left[\varepsilon\varepsilon^T\right]\mathbf{V}^T = \mathbf{V}\sigma^2\mathbf{\Omega}\mathbf{V}^T = \\ &= \sigma^2\mathbf{V}\left(\mathbf{V}^T\mathbf{V}\right)^{-1}\mathbf{V}^T = \sigma^2\mathbf{V}\mathbf{V}^{-1}\left(\mathbf{V}^T\right)^{-1}\mathbf{V}^T = \sigma^2\mathbf{I} \end{aligned}$$

Generalised Least Squares estimator (1)

- $\mathbf{y} = \mathbf{X}\beta + \varepsilon \qquad \varepsilon \sim \left(E[\varepsilon] = \mathbf{0}, E\left[\varepsilon\varepsilon^T\right] = \sigma^2\mathbf{\Omega}\right)$

With well-defined, symmetric matrix $\mathbf{\Omega}$...

...there is a unique decomposition $\mathbf{\Omega}^{-1} = \mathbf{V}^T\mathbf{V}$.

- Pre-multiply both sides of equation with $\mathbf{V}$:
  $\mathbf{V}\mathbf{y} = \mathbf{V}\mathbf{X}\beta + \mathbf{V}\varepsilon$
- The error term in this model is spherical:

$$\begin{aligned} var(\mathbf{V}\varepsilon) &= E\left[(\mathbf{V}\varepsilon)(\mathbf{V}\varepsilon)^T\right] = E\left[\mathbf{V}\varepsilon\varepsilon^T\mathbf{V}^T\right] = \\ &= \mathbf{V}E\left[\varepsilon\varepsilon^T\right]\mathbf{V}^T = \mathbf{V}\sigma^2\mathbf{\Omega}\mathbf{V}^T = \\ &= \sigma^2\mathbf{V}\left(\mathbf{V}^T\mathbf{V}\right)^{-1}\mathbf{V}^T = \sigma^2\mathbf{V}\mathbf{V}^{-1}\left(\mathbf{V}^T\right)^{-1}\mathbf{V}^T = \sigma^2\mathbf{I} \end{aligned}$$

Generalised Least Squares estimator (1)

- $\mathbf{y} = \mathbf{X}\beta + \varepsilon \qquad \varepsilon \sim \left(E\left[\varepsilon\right] = \mathbf{0}, E\left[\varepsilon\varepsilon^T\right] = \sigma^2\mathbf{\Omega}\right)$

With well-defined, symmetric matrix $\mathbf{\Omega}$...

...there is a unique decomposition $\mathbf{\Omega}^{-1} = \mathbf{V}^T\mathbf{V}$.

- Pre-multiply both sides of equation with $\mathbf{V}$:
  $\mathbf{V}\mathbf{y} = \mathbf{V}\mathbf{X}\beta + \mathbf{V}\varepsilon$
- The error term in this model is spherical:

$$
\begin{aligned}
var\left(\mathbf{V}\varepsilon\right) &= E\left[\left(\mathbf{V}\varepsilon\right)\left(\mathbf{V}\varepsilon\right)^T\right] = E\left[\mathbf{V}\varepsilon\varepsilon^T\mathbf{V}^T\right] = \\
&= \mathbf{V}E\left[\varepsilon\varepsilon^T\right]\mathbf{V}^T = \mathbf{V}\sigma^2\mathbf{\Omega}\mathbf{V}^T = \\
&= \sigma^2\mathbf{V}\left(\mathbf{V}^T\mathbf{V}\right)^{-1}\mathbf{V}^T = \sigma^2\mathbf{V}\mathbf{V}^{-1}\left(\mathbf{V}^T\right)^{-1}\mathbf{V}^T = \sigma^2\mathbf{I}
\end{aligned}
$$

GLS estimator (2)

$$\mathbf{Vy} = \mathbf{VX}\beta + \mathbf{V}\varepsilon$$

Assume that $\mathbf{V}$ (and, hence, $\mathbf{\Omega}$) is known. Then the estimation of the original equation with GLS is equivalent to the estimation of the above equation with OLS (using transformed data):

$$\hat{\beta}^{OLS} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}$$

$$\begin{aligned}\hat{\beta}^{GLS} &= \left[(\mathbf{VX})^T(\mathbf{VX})\right]^{-1}(\mathbf{VX})^T(\mathbf{Vy}) = \\ &= \left[\mathbf{X}^T\mathbf{V}^T\mathbf{VX}\right]^{-1}\mathbf{X}^T\mathbf{V}^T\mathbf{Vy} = \left[\mathbf{X}^T\mathbf{\Omega}^{-1}\mathbf{X}\right]^{-1}\mathbf{X}^T\mathbf{\Omega}^{-1}\mathbf{y}\end{aligned}$$

How to find matrix $\boldsymbol{\Omega}$?

- unknown and cannot be estimated in general (contains T^2 parameters under T observations)
- but can be expressed as a function of a low number of parameters
- this requires some assumption about the stochastic process for ε_t, e.g.:
  - $\varepsilon_t = \rho\varepsilon_{t-1} + \eta_t \qquad \eta \sim (\mathbf{0}, \sigma^2\mathbf{I})$, then $\hat{\Omega} = \Omega(\hat{\rho})$
  - $\varepsilon_t = \rho_1\varepsilon_{t-1} + \rho_2\varepsilon_{t-2} + \eta_t \qquad \eta \sim (\mathbf{0}, \sigma^2\mathbf{I})$, then $\hat{\Omega} = \Omega(\hat{\rho}_1, \hat{\rho}_2)$
  - etc.

How to find matrix $\mathbf{\Omega}$?

- unknown and cannot be estimated in general (contains T^2 parameters under T observations)
- but can be expressed as a function of a low number of parameters
- this requires some assumption about the stochastic process for ε_t, e.g.:
  - $\varepsilon_t = \rho\varepsilon_{t-1} + \eta_t \quad \eta \sim (\mathbf{0}, \sigma^2\mathbf{I})$, then $\hat{\Omega} = \Omega(\hat{\rho})$
  - $\varepsilon_t = \rho_1\varepsilon_{t-1} + \rho_2\varepsilon_{t-2} + \eta_t \quad \eta \sim (\mathbf{0}, \sigma^2\mathbf{I})$, then $\hat{\Omega} = \Omega(\hat{\rho}_1, \hat{\rho}_2)$
  - etc.

How to find matrix $\mathbf{\Omega}$?

- unknown and cannot be estimated in general (contains T^2 parameters under T observations)
- but can be expressed as a function of a low number of parameters
- this requires some assumption about the stochastic process for ε_t, e.g.:
  - $\varepsilon_t = \rho\varepsilon_{t-1} + \eta_t \qquad \eta \sim \left(\mathbf{0}, \sigma^2\mathbf{I}\right)$, then $\hat{\Omega} = \Omega\left(\hat{\rho}\right)$
  - $\varepsilon_t = \rho_1\varepsilon_{t-1} + \rho_2\varepsilon_{t-2} + \eta_t \qquad \eta \sim \left(\mathbf{0}, \sigma^2\mathbf{I}\right)$, then $\hat{\Omega} = \Omega\left(\hat{\rho}_1, \hat{\rho}_2\right)$
  - etc.

AR(1) example: $\hat{\Omega} = \Omega(\hat{\rho})$ (1)

Suppose: $\varepsilon_t = \rho\varepsilon_{t-1} + \eta_t \qquad \eta \sim \left(\mathbf{0}, \sigma_\eta^2 \mathbf{I}\right)$

$$var(\varepsilon_t) = var(\rho\varepsilon_{t-1} + \eta_t) = var(\rho^2\varepsilon_{t-2} + \rho\eta_{t-1} + \eta_t) =$$

$$var\left(\eta_t + \rho\eta_{t-1} + \rho^2\eta_{t-2} + \ldots\right) = \sigma_\eta^2 + \sigma_\eta^2\rho^2 + \sigma_\eta^2\rho^4 + \ldots = \frac{\sigma_\eta^2}{1-\rho^2}$$

$$cov(\varepsilon_t, \varepsilon_{t-1}) = cov\left(\rho\varepsilon_{t-1} + \eta_t, \varepsilon_{t-1}\right) = \rho var(\varepsilon_{t-1}) = \rho\frac{\sigma_\eta^2}{1-\rho^2}$$

$$cov(\varepsilon_t, \varepsilon_{t-2}) = cov\left(\rho^2\varepsilon_{t-2} + \rho\varepsilon_{t-1} + \eta_t, \varepsilon_{t-2}\right) = \rho^2 var(\varepsilon_{t-2}) = \rho^2\frac{\sigma_\eta^2}{1-\rho^2}$$

$\vdots$

$$\hat{\mathbf{\Omega}} = \frac{\sigma_\eta^2}{1-\rho^2}\begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{T-1} \\ \rho & 1 & \rho & \ddots & \vdots \\ \rho^2 & \rho & 1 & \ddots & \rho^2 \\ \vdots & \ddots & \ddots & \ddots & \rho \\ \rho^{T-1} & \cdots & \rho^2 & \rho & 1 \end{bmatrix}$$

AR(1) example: $\hat{\Omega} = \Omega(\hat{\rho})$ (2)

Under such circumstances, one can show that:

$$\hat{\boldsymbol{\Omega}}^{-1} = \frac{1}{\sigma_\eta^2}\begin{bmatrix} 1 & -\rho & 0 & \cdots & 0 \\ -\rho & 1+\rho^2 & -\rho & \ddots & \vdots \\ 0 & -\rho & 1+\rho^2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -\rho \\ 0 & \cdots & 0 & -\rho & 1 \end{bmatrix}$$

$$\boldsymbol{V} = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & 0 & \cdots & 0 \\ -\rho & 1 & 0 & \ddots & \vdots \\ 0 & -\rho & 1 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -\rho & 1 \end{bmatrix}$$

Proof: Compute $\hat{\boldsymbol{\Omega}}\hat{\boldsymbol{\Omega}}^{-1}$ (should be equal to $\boldsymbol{I}$) and $\boldsymbol{V}^T\boldsymbol{V}$ (should be equal to $\hat{\boldsymbol{\Omega}}^{-1}$).

"Feasible" GLS in practice

$$\hat{\Omega}$$

$$\downarrow$$

$$\hat{\beta}^{GLS}$$

"Feasible" GLS in practice

$$\hat{\Omega} \quad \leftarrow \quad \hat{\rho}$$

$$\downarrow$$

$$\hat{\beta}^{GLS}$$

"Feasible" GLS in practice

$$\begin{array}{ccc} \hat{\Omega} & \leftarrow & \hat{\rho} \\ & & \color{red}{\uparrow} \\ \downarrow & & \color{red}{\hat{\varepsilon}} \\ \hat{\beta}^{GLS} & & \end{array}$$

"Feasible" GLS in practice

$$\begin{array}{ccc}
\hat{\Omega} & \leftarrow & \hat{\rho} \\
 & & \uparrow \\
\downarrow & & \hat{\varepsilon} \\
 & & \color{red}{\uparrow} \\
\hat{\beta}^{GLS} & \color{red}{\rightarrow} & \color{red}{\hat{y}}
\end{array}$$

"Feasible" GLS in practice

$$
\begin{array}{ccccc}
 & & \hat{\Omega} & \leftarrow & \hat{\rho} \\
 & & & & \uparrow \\
 & & \downarrow & & \hat{\varepsilon} \\
\text{Start:} & & & & \uparrow \\
\hat{\beta}^{OLS} & \rightarrow & \hat{\beta}^{GLS} & \rightarrow & \hat{y}
\end{array}
$$

Cochrane-Orcutt's method

Special case of GLS: 1st order serial correlation assumed.

$$y_t = \mathbf{x}_t\boldsymbol{\beta} + \varepsilon_t \qquad \varepsilon_t = \rho_1\varepsilon_{t-1} + \eta_t$$

$$y_{t-1} = \mathbf{x}_{t-1}\boldsymbol{\beta} + \varepsilon_{t-1} \qquad \eta_t = \varepsilon_t - \rho_1\varepsilon_{t-1}$$

Subtract both sides with second equation multiplied by ρ_1:

$$y_t - \rho_1 y_{t-1} = \mathbf{x}_t\boldsymbol{\beta} - \rho_1\mathbf{x}_{t-1}\boldsymbol{\beta} + \varepsilon_t - \rho_1\varepsilon_t$$

$$\underbrace{y_t - \rho_1 y_{t-1}}_{\textit{transformed independent}} = \underbrace{(\mathbf{x}_t - \rho_1\mathbf{x}_{t-1})}_{\textit{transformed dependent}}\boldsymbol{\beta} + \underbrace{\eta_t}_{\textit{spherical error}}$$

Generalized Cochrane-Orcutt: serial correlation of higher orders.

Exercise (3/3)

Gasoline demand model

Look at our model again and compare the estimates (and their significance) obtained with.

1. OLS
2. OLS with robust standard errors
3. Cochrane-Orcutt method

Discuss.

Readings

- Greene: chapter "Serial correlation"
- Greene: chapter "Nonspherical Disturbances – The Generalized Regression Model" (sub-chapter about feasible GLS)

Homework

Gasoline demand model

Write manually an R programme performing the GLS estimation with AR(p) residuals, for any sensible value of p. Use it to estimate the version of the model with AR(2) residuals.

Other models

Do the exercises specified in the comments of the R code.