Lecture 3: Time series regression: Serial Correlation and Heteroskedasticity

Yu Bai

City University of Macau

Properties of OLS with Serially Correlated Errors

Unbiasedness and Consistency

- Notice that, assumption on $\text{Cov}\left(u_t, u_s \middle| \cdots\right)$ are not imposed when deriving unbiasedness and consistency.
- This means that even Assumptions TS.5(5') is violated, OLS estimator is still consistent (and also unbiased under strict exogeneity).

Efficiency and Inference

- Consider a simple regression model with AR(1) innovations:

$$y_t = \beta_1 x_t + u_t, \quad u_t = \rho u_{t-1} + e_t, \ \ t = 1, 2, \cdots, n.$$

- The OLS estimator $\hat{\beta}_{1n}$ can be written as

$$\hat{\beta}_{1n} = \beta_1 + \frac{\sum_{t=1}^{n} x_t u_t}{\sum_{t=1}^{n} x_t^2}.$$

- The variance of $\hat{\beta}_{1n}$ (conditional on $\mathbf{X}$) takes the following form

$$\mathbb{V}\left(\hat{\beta}_{1n}\right) = \frac{\sigma^2}{\sum_{t=1}^{n} x_t^2} + 2\left(\sigma^2 / \left(\sum_{t=1}^{n} x_t^2\right)^2\right) \sum_{t=1}^{n-1} \sum_{j=1}^{n-t} \rho^j x_t x_{t+j},$$

where $\sigma^2 = \mathbb{V}(u_t)$ and we have used the fact that $\mathbb{E}(u_t u_{t+j}) = \rho^j \sigma^2$.

Efficiency and Inference

- If we ignore the serial correlation and estimate the variance in the usual way, the variance estimator will usually be biased.
- In most applications, the neglected part is positive, the usual OLS variance formula *understates* the true variance of the OLS estimator.
- t-statistics are no longer valid for testing single hypotheses. The usual F and LM-statistics for testing multiple hypotheses are also invalid.

Going further questions

Suppose that, rather than AR(1) model, u_t follows the MA(1) model: $u_t = e_t + \alpha e_{t-1}$. Find $\mathbb{V}\left(\hat{\beta}_{1n}\right)$.

Goodness of Fit

- R-squared in population:

$$R^2 = 1 - \frac{\sigma_u^2}{\sigma_y^2}$$

- When the data are stationary and weakly dependent, the WLLN kicks in, so that

$$\frac{1}{n}\sum_{t=1}^{n}\left(y_t - \mathbf{X}_t'\hat{\boldsymbol{\beta}}_n\right)^2 \xrightarrow{p} \sigma_u^2, \quad \frac{1}{n}\sum_{t=1}^{n}(y_t - \overline{y}_n)^2 \xrightarrow{p} \sigma_y^2.$$

- Nothing to worry about...

Serial Correlation in the Presence of Lagged Dependent Variables

Statement

OLS is inconsistent in the presence of lagged dependent variables and serially correlated errors.

Serial Correlation in the Presence of Lagged Dependent Variables

Statement

OLS is inconsistent in the presence of lagged dependent variables and serially correlated errors.

- Unfortunately, as a general assertion, this statement is false. There is a version of the statement that is correct, but it is important to be very precise.

- Consider an AR(1) model with $|\beta_1| < 1$:

$$y_t = \beta_1 y_{t-1} + u_t,$$

where $\mathbb{E}\left(u_t | y_{t-1}\right) = 0$. Then, we know that OLS estimator is consistent:

$$\hat{\beta}_{1n} = \frac{\sum_{t=1}^{n} y_{t-1} y_t}{\sum_{t=1}^{n} y_{t-1}^2} \xrightarrow{p} \beta_1$$

- Now suppose that $(y_t)_t$ follows an ARMA(1,1) process:

$$y_t = \beta_1 y_{t-1} + u_t + \theta_1 u_{t-1},$$

but we still estimate β_1 using OLS. Is it consistent?

- Consider an AR(1) model with $|\beta_1| < 1$:

$$y_t = \beta_1 y_{t-1} + u_t,$$

where $\mathbb{E}\left(u_t | y_{t-1}\right) = 0$. Then, we know that OLS estimator is consistent:

$$\hat{\beta}_{1n} = \frac{\sum_{t=1}^{n} y_{t-1} y_t}{\sum_{t=1}^{n} y_{t-1}^2} \xrightarrow{p} \beta_1$$

- Now suppose that $\left(y_t\right)_t$ follows an ARMA(1,1) process:

$$y_t = \beta_1 y_{t-1} + u_t + \theta_1 u_{t-1},$$

but we still estimate β_1 using OLS. Is it consistent?

- Some derivations:

$$\hat{\beta}_{1n} - \beta_1 = \frac{\frac{1}{n}\sum_{t=1}^{n} y_{t-1} u_t}{\frac{1}{n}\sum_{t=1}^{n} y_{t-1}^2} + \frac{\theta_1 \frac{1}{n}\sum_{t=1}^{n} u_{t-1} u_t}{\frac{1}{n}\sum_{t=1}^{n} y_{t-1}^2}$$

- Why should we care this kind of things?
- Suppose that we want to forecast y_{T+1} using the model:

$$y_t = \beta_1 y_{t-1} + u_t,$$
$$u_t = \theta_1 u_{t-1} + e_t,$$

where $(e_t)_t$ satisfies $\mathbb{E}(e_t | e_{t-1}, e_{t-2}, \cdots) = 0$.

- It can be shown that the optimal forecast is given by

$$\mathbb{E}(y_{T+1} | \mathcal{F}_T) = \beta_1 y_T + \theta_1 (y_T - \beta_1 y_{T-1}).$$

- What does this imply?

Serial Correlation–Robust Inference after OLS

- For a linear regression model with k regressors:

$$y_t = \beta_0 + \sum_{d=1}^{k} \beta_d x_{td} + u_t, \quad t = 1, 2, \cdots, n$$

- Write x_{t1} as a linear function of the remaining independent variables and an error term

$$x_{t1} = \delta_0 + \sum_{d=2}^{k} \delta_d x_{td} + r_t,$$

where the error r_t has zero mean and is uncorrelated with $x_{t2}, x_{t3}, \cdots, x_{tk}$.

- Wooldridge (1989, EL) has shown that the asymptotic variance of the OLS estimator $\hat{\beta}_{1n}$ is given by

$$\left(\sum_{t=1}^{n}\mathbb{E}(r_t^2)\right)^{-2}\mathbb{V}\left(\sum_{t=1}^{n}r_t u_t\right).$$

- If Assumption TS.5' fails,

$$\mathbb{V}\left(\sum_{t=1}^{n}r_t u_t\right)=\sum_{t=1}^{n}\mathbb{V}(r_t^2 u_t^2)+2\sum_{t=1}^{n}\sum_{s=1}^{t-1}\mathrm{cov}\left(r_t u_t r_s u_s\right),$$

where the terms in red is nonzero.

- Needs truncation to make it practically feasible and replace u_t with $\hat{u}_t = y_t - \mathbf{X}_t'\hat{\beta}_n$, but, say, can we use formula as below?

$$\frac{1}{n}\sum_{t=1}^{n}\hat{r}_t^2\hat{u}_t^2+2\frac{1}{n}\sum_{t=1}^{n}\hat{r}_t\hat{u}_t\hat{r}_{t-1}\hat{u}_{t-1}+\cdots$$

HAC standard errors

- For a chosen integer $g > 0$, define

$$\hat{v} = \sum_{t=1}^{n} \hat{a}_t^2 + 2 \sum_{h=1}^{g} \left[1 - \frac{h}{g+1}\right] \left(\sum_{t=h+1}^{n} \hat{a}_t \hat{a}_{t-h} \right),$$

  where $\hat{a}_t = \hat{r}_t \hat{u}_t$.

- The weights $\omega_h = 1 - h/(g+1)$ is first suggested in Newey and West (1987) so the above is also called **Newey-West standard errors**.
- Let "se $\left(\hat{\beta}_{1n}\right)$" be the usual (but incorrect) standard error and $\hat{\sigma}$ be the usual root mean squared error from estimating the regression, the **heteroskedastic and autocorrelation consistent (HAC)** standard error is simply

$$\text{se}\left(\hat{\beta}_{1n}\right) = \left[\text{"se}\left(\hat{\beta}_{1n}\right)\text{"}/\hat{\sigma}\right]^2 \sqrt{\hat{v}}.$$

How to choose g?

- The integer g is often called **truncation lag**.
- Consistency requires that $g := g_n \to \infty$ but how should we choose g in practice?
- Should be dependent on data frequency
- Can set $g_n = cn^k$ and choose c based on rule-of-thumb
  - $k = 1/4$ as recommended in Stock and Watson (2014)'s textbook, as consistency requires g_n grows at a slower rate than $n^{1/4}$
- Can be easily implemented in software by adding options

Some more discussions...

- Empirically, the **HAC** standard errors are typically larger than the usual OLS standard errors when there is serial correlation.
- Might be sensitive to the choice of g, since
  - Practical recommendations are based on asymptotic argument.
- How about a fixed g?
  - **heteroskedastic and autocorrelation robust (HAR)** standard error

Testing for Serial Correlation

Testing for Serial Correlation with Strictly Exogenous Regressors

- Consider the model
$$y_t = \mathbf{X}_t'\boldsymbol{\beta} + u_t, \quad t = 1, 2, \cdots, n,$$
where $\mathbb{E}\left(u_t|\mathbf{X}\right) = 0$.
- We would like to test whether there is a first-order serial correlation:
$$u_t = \rho u_{t-1} + e_t.$$
- The null hypothesis is simply $\mathcal{H}_0 : \rho = 0$.
- What should we do?

Testing for Serial Correlation with Strictly Exogenous Regressors

We can summarize the asymptotic test for AR(1) serial correlation as below.

1. Run the OLS regression to obtain the OLS residuals, $\hat{u}_t$, for all $t = 1, 2, \cdots, n$.
2. Run the regression of $\hat{u}_t$ on $\hat{u}_{t-1}$, for all $t = 2, \cdots, n$, obtaining the coefficient $\hat{\rho}_n$ on $\hat{u}_{t-1}$ and its t-statistic $t_{\hat{\rho}_n}$.
3. Use $t_{\hat{\rho}_n}$ to test $\mathcal{H}_0 : \rho = 0$.

Testing for Serial Correlation without Strictly Exogenous Regressors

Just change the second step:

1. Run the OLS regression to obtain the OLS residuals, $\hat{u}_t$, for all $t = 1, 2, \cdots, n$.
2. Run the regression of $\hat{u}_t$ on $\hat{u}_{t-1}$, 1, $x_{t1}, \cdots, x_{tk}$, for all $t = 2, \cdots, n$, obtaining the coefficient $\hat{\rho}_n$ on $\hat{u}_{t-1}$ and its t-statistic $t_{\hat{\rho}_n}$.
3. Use $t_{\hat{\rho}_n}$ to test $\mathcal{H}_0 : \rho = 0$.

A few reminders:

- Should have an intercept in the second step (according to the book)
- Why do we need to add variables in the second step?
- Can easily adjust for heteroscedasticity and serial correlation when computing $\text{se}(\hat{\rho}_n)$

Testing for Higher-Order Serial Correlation

- Can easily adapted by adding more lags in the second step and use F-test for joint significance
- Alternatively, we can use the Lagrange multiplier (LM) test:

$$\mathrm{LM}_n = (n-q)R^2_{\hat{u}} \overset{a}{\sim} \chi^2_q,$$

  where $R^2_{\hat{u}}$ is the usual R-squared from the regression of $\hat{u}_t$ on $\hat{u}_{t-1}, \cdots, \hat{u}_{t-q}$, $x_{t1}, \cdots, x_{tk}$.
- This is also called the **Breusch-Godfrey test** for AR(q) serial correlation.

- With quarterly or monthly data that have not been seasonally adjusted, we sometimes wish to test for seasonal forms of serial correlation.
- This is left as an exercise.

Going further questions

Suppose you have quarterly data and you want to test for the presence of first-order or fourth-order serial correlation. With strictly exogenous regressors, how would you proceed?

Correcting for Serial Correlation with Strictly Exogenous Regressors

Some motivation...

- We have learned how to use **HAC** to have robust standard errors in the presence of heteroscedasticity and serial correlation.
- This is a nonparametric approach. We do not specify the form of heteroscedasticity and serial correlation. However,
  - We need to specify the truncation lag parameter g. Results might be sensitive to the choice of g.
- How about a parametric approach?

Obtaining the Best Linear Unbiased Estimator in the AR(1) Model

- Consider the model with a single explanatory variable:

$$y_t = \beta_0 + \beta_1 x_t + u_t, \quad t = 1, 2, \cdots, n.$$

- We do not assume no serial correlation, but specify an AR(1) dynamics for u_t:

$$u_t = \rho u_{t-1} + e_t,$$

where $|\rho| < 1$ and $(e_t)_t$ is $i.i.d.$ with zero mean and variance σ_e^2.

- By quasi-differencing, we obtain

$$\tilde{y}_t = (1-\rho)\beta_0 + \beta_1 \tilde{x}_t + e_t,$$

  where $\tilde{y}_t = y_t - \rho y_{t-1}$ and $\tilde{x}_t = x_t - \rho x_{t-1}$.
- The error term e_t now satisfies Assumptions TS.4-5.
- However, the above is not defined for $t = 1$.
- We note that, for $t = 1$

$$(1-\rho^2)^{1/2} y_1 = (1-\rho^2)^{1/2}\beta_0 + \beta_1 (1-\rho^2)^{1/2} x_1 + (1-\rho^2)^{1/2} u_1,$$

  where $\mathbb{V}\left((1-\rho^2)^{1/2} u_1\right) = \sigma_e^2$.
- Applying OLS on the transformed model is essentially another example of a generalized least squares (or GLS) estimator, which we have seen in Chapter 8.

Going deeper...

How do we construct the OLS estimator based on the transformed model?

Feasible GLS Estimation

- Very simple, just replace ρ with $\hat{\rho}_n$ from OLS regression of $\hat{u}_t$ on $\hat{u}_{t-1}$.
- Such estimator is called
  - **Cochrane-Orcutt (CO) estimation** if the first observation is omitted;
  - **Prais-Winsten (PW) estimation** if the first observation is used.
- In practice, we may need iteration to improve accuracy.

Comparing OLS and FGLS

- It can be shown that, for the regression model with AR(1) error, consistency of FGLS requires u_t to be uncorrelated with x_{t-1}, x_t, and x_{t+1}.
- What should we do in practice?
  - When both yield similar estimates, ...
  - When there are practical differences, ...

Correcting for Higher-Order Serial Correlation

- Rather straightforward...
- May be more involved in dealing with initial observations...
  - Don't worry, we rarely have to compute these by ourselves.

What if the Serial Correlation Model Is Wrong?

George Box

All models are wrong, some are useful.

- Start with higher-order serial correlation?
- Using HAC after the initial AR(1) specification for the error term?

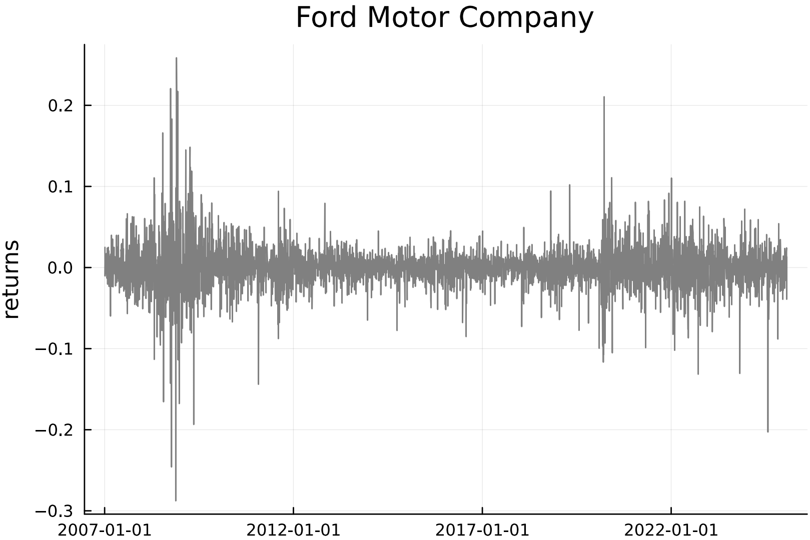
Keep in mind that serial correlation only causes standard formula for OLS variance to be invalid, nothing related to the consistency of $\hat{\beta}_n$.

Heteroskedasticity in Time Series Regressions

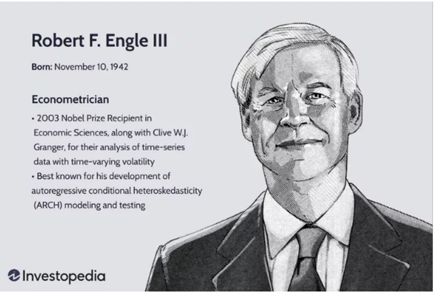
Testing for Heteroskedasticity

- To use what we have learned, we must assume that the error term should be uncorrelated.
- Serial correlation perhaps needs to come first.
- Breusch-Pagan test? White test? Weighted least squares?

Autoregressive Conditional Heteroskedasticity



Autoregressive Conditional Heteroskedasticity



Econometrica, Vol. 50, No. 4 (July, 1982)

AUTOREGRESSIVE CONDITIONAL HETEROSCEDASTICITY WITH ESTIMATES OF THE VARIANCE OF UNITED KINGDOM INFLATION[1]

BY ROBERT F. ENGLE

Traditional econometric models assume a constant one-period forecast variance. To generalize this implausible assumption, a new class of stochastic processes called autoregressive conditional heteroscedastic (ARCH) processes are introduced in this paper. These are mean zero, serially uncorrelated processes with nonconstant variances conditional on the past, but constant unconditional variances. For such processes, the recent past gives information about the one-period forecast variance.

Autoregressive Conditional Heteroskedasticity

- Suppose that we have the model

$$r_t = \sqrt{h_t}\epsilon_t, \;\; \epsilon_t \overset{i.i.d.}{\sim} (0, 1)$$

- How to model those volatility "clustering" effects?
- Specify an ARCH(1) model (Engle (1982, ECTA)):

$$h_t = \alpha_0 + \alpha_1 r_{t-1}^2$$

- It is like an AR(1) for variance, so we must have restrictions on α_0, α_1.

How to estimate the model parameters?

- Notice that
$$r_t^2 = h_t + h_t\epsilon_t^2 - h_t = \alpha_0 + \alpha_1 r_{t-1}^2 + u_t,$$
where $u_t = h_t\left(\epsilon^2 - 1\right)$.
- This suggests that $(\alpha_0, \alpha_1)'$ can be simply estimated using OLS!
- Unlikely to be efficient and requires high order moment restrictions on ϵ_t.

Regression models with ARCH errors

- Suppose that we add the mean terms for systematic risk:

$$r_{it} = \alpha_i + \beta_i r_{M,t} + \sqrt{h_{it}}\epsilon_{it}, \;\; \epsilon_{it} \overset{i.i.d.}{\sim} (0,1)$$
$$h_{it} = \alpha_{0i} + \alpha_{1i} r_{i,t-1}^2.$$

- It can be shown that the error terms $u_{it} = \sqrt{h_{it}}\epsilon_{it}$ have zero mean, variance $\frac{\alpha_{0i}}{1-\alpha_{1i}}$, and are serially uncorrelated.
- OLS should still have desirable properties with ARCH errors.

Question

If so, why should we still care about ARCH errors?

- Can also be extended to dynamic models...

Problem 1

Consider the model:

$$\begin{aligned} y_t &= \beta_0 + \sum_{k=1}^{K} \beta_k x_{tk} + u_t, \\ u_t &= \sqrt{h_t} v_t, \\ v_t &= \rho v_{t-1} + e_t, \ \ |\rho| < 1, \end{aligned}$$

where the explanatory variables $\mathbf{X}$ are independent of e_t for all t, and h_t is a function of the x_{tj}. The process $(e_t)_t$ has zero mean and constant variance σ_e^2, and is serially uncorrelated. Briefly describe the FGLS estimation procedure for $(\beta_0, \beta_1, \cdots, \beta_K)'$.

Problem 2

True or False?

1. The Cochrane-Orcutt and Prais-Winsten methods are both used to obtain valid standard errors for the OLS estimates when there is a serial correlation.
2. If the errors in a regression model contain ARCH, they must be serially correlated.