AFRE 835: Introductory Econometrics

Chapter 12: Serial Correlation and Heteroskedasticity in Time Series Regressions

Spring 2017

Introduction

- Wooldridge argues at the end of chapter 10, that a dynamically complete model should not have serially correlated errors.

  ...so that one can interpret serial correlation in the errors as an indication that the model is not dynamically correct.
- This chapter focuses on what one can do about
  - testing for serial correlation;
  - adjust for it when it is found.

Outline

Properties of OLS with Serially Correlated Errors

- Under strict exogeneity (TS.3), as well as TS.1 and TS.2, the OLS estimator in the time series setting is unbiased and consistent (Theorem 10.1).

  ... Serial correlation is not ruled out by these assumptions.
- Under contemporaneous exogeneity (TS.3'), as well as TS.2' and weak dependence and stationarity (TS.1'), the OLS estimators is consistent (though not necessarily unbiased - Theorem 11.1).

  ... Again, serial correlation is not ruled out by these assumptions.
- However, the presence of serial correlation does impact the efficiency of and variance estimators for the OLS estimators

The Variance of $\hat{\beta}_1$ in the Simple Regression Model

- Recall that in the simple regression model

$$\hat{\beta}_1 = \beta_1 + SST_x^{-1} \sum_{t=1}^{n} \ddot{x}_t u_t \tag{1}$$

where $\ddot{x}_t \equiv x_t - \bar{x}$ and $SST_x = \sum_{t=1}^{n} \ddot{x}_t^2$.

- Then, given homoskedasticity

$$\begin{aligned}
Var(\hat{\beta}_1|\boldsymbol{X}) &= SST_x^{-2} Var\left[\sum_{t=1}^{n} \ddot{x}_t u_t|\boldsymbol{X}\right] \\
&= SST_x^{-2}\left[\sum_{t=1}^{n} Var(\ddot{x}_t u_t|\boldsymbol{X}) + 2\sum_{t=1}^{n}\sum_{j=1}^{n-t} Cov(\ddot{x}_t u_t, \ddot{x}_{t+j} u_{t+j}|\boldsymbol{X})\right] \\
&= SST_x^{-2}\left[\sum_{t=1}^{n} \ddot{x}_t^2 Var(u_t|\boldsymbol{X})\right] + 2SST_x^{-2}\left[\sum_{t=1}^{n}\sum_{j=1}^{n-t} \ddot{x}_t \ddot{x}_{t+j} Cov(u_t, u_{t+j}|\boldsymbol{X})\right] \\
&= \frac{\sigma^2}{SST_x} + \frac{2\sigma^2}{SST_x^2}\left[\sum_{t=1}^{n}\sum_{j=1}^{n-t} \ddot{x}_t \ddot{x}_{t+j} \rho_j\right] \qquad \text{where } \rho_j = Corr(u_t, u_{t+j})
\end{aligned}$$

The AR(1) Case

- If the error terms follow an AR(1) process, then

$$u_t = \rho u_{t-1} + e_t, \quad t = 1, 2, \ldots \tag{2}$$

with $|\rho| < 1$ and the e_t's are uncorrelated with mean zero and variance σ^2.

- In this case, $\rho_j = \rho^j \;\; \forall j$ and the OLS estimator's variance reduces to

$$Var(\hat{\beta}_1|\boldsymbol{X}) = \frac{\sigma^2}{SST_x} + \frac{2\sigma^2}{SST_x^2}\left[\sum_{t=1}^{n}\sum_{j=1}^{n-t} \ddot{x}_t \ddot{x}_{t+j} \rho^j\right]$$

- The first term is the usual OLS variance when $\rho = 0$.
- When there is serial correlation, the second term will *typically* be positive, so the usual OLS variance will be biased downward (i.e., too small).

Serial Correlation and Lagged Dependent Variables

- It is often claimed that OLS is inconsistent if one has both lagged dependent variables and serial correlation.
- While this combination *can* be a problem, it need not be.
- Wooldridge (p. 415) gives the counter-example where

$$y_t = \beta_0 + \beta_1 y_{t-1} + u_t \tag{3}$$

  with the zero conditional mean assumption TS.3' (contemporaneous exogeneity) is satisfied; i.e.,

$$E(u_t | y_{t-1}) = 0. \tag{4}$$

- Based on Theorem 11.1, OLS is consistent in this setting, but we can still have serial correlation (with strict exogeneity violated).
- For example, if $Cov(u_t, y_{t-2}) \neq 0$, then

$$\begin{aligned} Cov(u_t, u_{t-1}) &= E\left[u_t(y_{t-1} - \beta_0 - \beta_1 y_{t-2})\right] \\ &= \beta_1 E\left[u_t y_{t-2}\right] \neq 0. \end{aligned} \tag{5}$$

Serial Correlation and Lagged Dependent Variables (cont'd)

- However, one can specify a a specific form of serial correlation that will lead to a violation of TS.3'.
- If, for example, u_t follows an AR1 process, then

$$Cov(y_{t-1}, u_t) = E[y_{t-1}(\rho u_{t-1} + e_t)] = \rho Cov(y_{t-1} u_{t-1}) \neq 0 \quad \forall \rho \neq 0 \tag{6}$$

  ...which clearly violates the contemporaneous exogeneity assumption in Theorem 11.1.
- As Wooldridge notes, the AR1 model for the error term actually suggests that the model is not dynamically complete.

  In particular, it suggests that the right formulation for the model of y_t is an AR2 model.

Serial Correlation and Lagged Dependent Variables (cont'd)

- To see this, note that

$$\begin{aligned} y_t &= \beta_0 + \beta_1 y_{t-1} + u_t \\ &= \beta_0 + \beta_1 y_{t-1} + \rho u_{t-1} + e_t \\ &= \beta_0 + \beta_1 y_{t-1} + \rho(y_{t-1} - \beta_0 - \beta_1 y_{t-2}) + e_t \\ &= \beta_0 + (\beta_1 + \rho) y_{t-1} + (-\rho\beta_1) y_{t-2} + e_t \\ &= \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + e_t. \end{aligned} \tag{7}$$

Testing for AR1 Serial Correlation with Strict Exogeneity

- Suppose we a multiple linear regression model, with

$$y_t = \beta_0 + \beta_1 x_{t1} + \cdots + \beta_k x_{tk} + u_t \tag{8}$$

where $E(u_t|\boldsymbol{X}) = 0$.

- Let $u_t = \rho u_{t-1} + e_t$ and $|\rho| < 1$, where e_t's are uncorrelated with mean zero and variance σ^2.
- Finally, assume that

$$E(e_t|u_{t-1}, u_{t-2}, \ldots) = 0 \tag{9}$$

and

$$Var(e_t|u_{t-1}) = Var(e_t) = \sigma_e^2. \tag{10}$$

Testing for AR1 Serial Correlation with Strict Exogeneity (cont'd)

- Under the null hypothesis $H_0 : \rho = 0$, if we observed u_t, then we could estimate the AR1 model $u_t = \rho u_{t-1} + e_t$ and test the hypothesis.
- It turns out that the hypothesis test based on fitted residuals from OLS estimation of (8) is justified asymptotically.

  ... That is, one would
  - construct fitted residuals from OLS estimation of (8),
  - run the OLS regression $\hat{u}_t = \rho \hat{u}_{t-1} + error_t$, and
  - test the hypothesis that $\rho = 0$ in the usual way.
- Despite being based on an AR1 model, the test will typically detect other forms of serial correlation.
- The **Durbin-Watson (DW) statistic** provides an alternative test, but requires the full set of CLM assumptions (including normality) and often provides indeterminate results (See Wooldridge, pp 418-419).

Testing for AR1 Serial Correlation *without* Strict Exogeneity

- Without strict exogeneity, the t-statistic above needs modification.
- Durbin suggested the following approach:
  - construct fitted residuals from OLS estimation of (8),
  - run the OLS regression

    $$\hat{u}_t = \rho \hat{u}_{t-1} + \delta_0 + \delta_1 x_{t1} + \cdots + \delta_k x_{tk} + error_t \tag{11}$$

    and
  - test the hypothesis that $\rho = 0$ in the usual way.
- Note that (11) allows u_{t-1} to be correlated with the x_{tj}'s.
- One can also generalize the above to
  - make the test robust to heteroskedasticity
  - test for higher order serial correlation (including quarterly or monthly correlation patterns); e.g., Breusch-Godfrey LM test for AR(q).

Correcting for Serial Correlation

- With serial correlation (and strictly exogenous variables), we violate assumption TS.5, one of the Gauss-Markov assumptions.
- Without this assumption, the OLS estimator is no longer *BLUE*.
- We saw a similar problem in the cross-sectional setting, when heteroskedasticity led to OLS no longer being *BLUE*.

  ...In that case, WLS provided a way to transform the model and obtain a new *BLUE* estimator.
- A transformation also exists in for serial correlation in the time series setting.
- We'll consider in detail the case that the errors follow an AR1 process, but the idea generalizes.

Quasi-Differencing in the AR1 setting

- The transformation in the AR1 setting relies on knowing the exact form of the serial correlation; with $u_t = \rho u_{t-1} + e_t$.
- For periods t and $t-1$, and a single regressor, we can write

$$y_t = \beta_0 + \beta_1 x_t + u_t \tag{12}$$
$$\rho y_{t-1} = \rho\beta_0 + \beta_1 \rho x_{t-1} + \rho u_{t-1} \tag{13}$$

  where we have simply multiplied the second line by ρ.
- Subtracting equation (13) from (12) yields

$$\tilde{y}_t = \beta_0 \tilde{x}_{t0} + \beta_1 \tilde{x}_{t1} + \tilde{e}_t \quad t \geq 2 \tag{14}$$

  where $\tilde{y}_t \equiv y_t - \rho y_{t-1}$, $\tilde{x}_{t1} \equiv x_t - \rho x_{t-1}$, $\tilde{x}_{t0} \equiv (1-\rho)$, and $\tilde{e}_t \equiv u_t - \rho u_{t-1} = e_t$.
- The transformed model satisfies the assumptions of Gauss-Markov Theorem (10.4), but the OLS applied to it is not quite *BLUE* because we have lost one observation.

Quasi-Differencing in the AR1 setting (cont'd)

- We could consider just using the first observation without transforming it.
- However, though we would still not have serially correlated errors, we would have heteroskedasticity, since $Var(u_t) = \frac{\sigma_e^2}{1-\rho^2} > Var(e_t)$.
- This problem can be fixed by applying a form of WLS to the first observation, multiplying through by $\sqrt{1-\rho^2}$; i.e.,

$$\sqrt{1-\rho^2}y_1 = \beta_0\sqrt{1-\rho^2} + \beta_1\sqrt{1-\rho^2}x_1 + \sqrt{1-\rho^2}u_1$$

or

$$\tilde{y}_1 = \beta_0\tilde{x}_{10} + \beta_1\tilde{x}_{11} + \tilde{e}_1 \tag{15}$$

- Adding this transformed initial observation to (14), the resulting GLS estimator will be *BLUE*, since the transformed model satisfied the Gauss-Markov Theorem assumptions.

GLS More Generally

- Both WLS and the above translation to correct for serial correlation fall into a class of estimators known as **Generalized Least Squares** applied to a linear regression model of the form

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}. \tag{16}$$

- They both are addressing violations of the Gauss-Markov assumptions that imply $Var(\boldsymbol{u}|\boldsymbol{X}) = \sigma^2\boldsymbol{I_n}$ where $\boldsymbol{I_n}$ is an $n \times n$ identity matrix.
  - In the WLS case, the violation takes the form of diagonal elements that are not the same;
  - In the case of serial correlation, some off-diagonal elements are non-zero.
- If $Var(\boldsymbol{u}|\boldsymbol{X}) \neq \sigma^2\boldsymbol{I_n}$, then the OLS estimator is no longer *BLUE*.
- GLS fixes this problem by transforming the model to restore the Gauss-Markov conditions.

GLS (cont'd)

- Suppose that $Var(\boldsymbol{u}|\boldsymbol{X}) = \sigma^2\boldsymbol{\Omega}$, where $\boldsymbol{\Omega}$ is a symmetric positive definite matrix and $\boldsymbol{\Omega} \neq \boldsymbol{I}_n$.
- Then $\boldsymbol{\Omega}^{-1}$ is also positive definite and there exists a nonsingular matrix $\boldsymbol{P}$ such that $\boldsymbol{P}'\boldsymbol{P} = \boldsymbol{\Omega}^{-1}$
- If we pre-multiply our model in (16) by $\boldsymbol{P}$ we get

$$\boldsymbol{P}\boldsymbol{y} = \boldsymbol{P}\boldsymbol{X}\beta + \boldsymbol{P}\boldsymbol{u}$$

or

$$\tilde{\boldsymbol{y}} = \tilde{\boldsymbol{X}}\beta + \tilde{\boldsymbol{u}}. \tag{17}$$

where $\tilde{\boldsymbol{y}} = \boldsymbol{P}\boldsymbol{y}$, $\tilde{\boldsymbol{X}} = \boldsymbol{P}\boldsymbol{X}$, and $\tilde{\boldsymbol{u}} = \boldsymbol{P}\boldsymbol{u}$.

- The model in (17) satisfies the Gauss-Markov Theorem assumptions, since $Var(\tilde{\boldsymbol{u}}|\boldsymbol{X}) = Var(\boldsymbol{P}\boldsymbol{u}|\boldsymbol{X}) = \boldsymbol{P}E(\boldsymbol{u}\boldsymbol{u}'|\boldsymbol{X})\boldsymbol{P}' = \sigma^2\boldsymbol{P}\boldsymbol{\Omega}\boldsymbol{P}' = \sigma^2\boldsymbol{P}(\boldsymbol{P}'\boldsymbol{P})^{-1}\boldsymbol{P}' = \sigma^2\boldsymbol{I}_n$.

GLS Examples

- For the AR1 model we considered above:

$$\boldsymbol{\Omega} = \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \rho^2 & \cdots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \rho^{n-4} & \cdots & 1 \end{bmatrix} \tag{18}$$

- It turns out that $\boldsymbol{P}$ is given by

$$\boldsymbol{P} = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & 0 & 0 & \cdots & 0 \\ -\rho & 1 & 0 & 0 & \cdots & 0 \\ 0 & -\rho & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \tag{19}$$

...which is precisely the transformation outlined above.

Feasible GLS

- While the GLS estimator is appealing, it requires knowing $\boldsymbol{\Omega}$.
- For the AR1 model, it requires ρ.
- The **Feasible GLS (FGLS) estimator** replaces $\boldsymbol{\Omega}$ with a consistent estimator of it.
- For the AR1 model, $\hat{\rho}$ is obtained by
  - first regressing y_t on $\boldsymbol{x}_t$ and obtaining fitted residuals $\hat{u}_t$.
  - regressing $\hat{u}_t$ on $\hat{u}_{t-1}$ to obtain $\hat{\rho}$.
  - applying OLS to the transformed model either with or without the first observation.

    ...The former is called the **Prais-Winsten estimator**, while the latter is called the **Cochrane-Orcutt estimator**.
- Unfortunately, the FGLS estimator also requires additional assumptions (e.g., strict exogeneity) in order to insure consistency. (See Wooldridge p. 427)
- One can generalize the above estimator to allow for higher order serial correlation; e.g., AR(q).

Choosing Between OLS and FGLS

- It can be challenging choosing between OLS and FGLS.
- OLS is consistent under weaker conditions generally.
- FGLS is more efficient if the error structure is correctly specified *and* the additional required assumptions are true.
- Even if not precisely correct, FGLS can reduce the departures from the Gauss-Markov conditions and *approximately* eliminate unit roots.

Differencing and Serial Correlation

- As we saw in chapter 11, applying OLS to highly persistent data (e.g., variables with unit roots) can be misleading.
- First differencing can eliminate the problem, as in the case of a random walk.
- First differencing can also *approximately* fix the problems associated with model exhibiting a high degree of serial correlation.
- This would be the case, for example, if the error terms u_t follows an AR1 process with ρ close to, but still less than, one.

Serial Correlation-Robust Inference after OLS

- As noted above, FGLS requires additional assumptions (e.g., strict exogeneity).
- If these assumptions are violated, FGLS may not even be consistent, let alone efficient.
- An alternative is to stick with OLS and correct the standard errors for fairly arbitrary forms of serial correlation.
- The so-called **serial correlation-robust standard errors**, which are also robust to heteroskedasticity, do precisely this.
- See section 12.5 in Wooldridge.

Heteroskedasticity in Time Series Regressions

- One can also correct for heteroskedasticity in the context of time series models, though this is rarely the focus.
- There are also forms of heteroskedasticity that have dynamic components.
- The first order **autoregressive conditional heteroskedasticity (ARCH)** model, for example, assumes that

$$E(u_t^2|u_{t-1}) = \alpha_0 + \alpha_1 u_{t-1}^2 \tag{20}$$

- These models have been used extensively in the empirical finance literature to understand market volatility.