Lecture 08

Resuming...

With time series, two problems appear:

- Correlation among errors: $\operatorname{corr}(u_t u_s) \neq 0, \quad t \neq s$.
    - This is usually an inefficiency problem.
- Correlation of RHS variables with error terms.
    - This is usually a bias/inconsistency problem.

The model:

$$y_t = \beta_1 + \beta_2 x_{2t} + u_t, \quad u_t = \rho u_{t-1} + e_t$$

features correlation among the errors, when $\rho \neq 0$.

- This incurs a loss of efficiency.
    - In this case, inefficiency expresses itself through small standard errors.
        - The results are very large t and F statistics.
        - It might give the impression of statistical significance, when there is actually not.
    - Then again, it is still unbiased and consistent.
- Note that in this case x_{2t} is not a random variable, it is given.
    - Since there is no randomness in x_{2t}, it cannot be correlated with the randomness if u_t.

For the model:

$$y_t = \beta_1 + \beta_2 x_{2t} + u_t, \quad x_{2t} = y_{t-1}$$

- There is no correlation among the errors: $u_t \perp u_s, \; \forall \, t \neq s$.
- We know that $\hat{\beta}_2^{\text{OLS}} = \beta_2 + \dfrac{\sum(x_{2t} - \bar{x}_2)u_t}{\sum(x_{2t} - \bar{x}_2)^2}$.
    - $a_t := \dfrac{x_{2t} - \bar{x}_2}{\sum(x_{2t} - \bar{x}_2)^2} \quad \Rightarrow \quad \hat{\beta}_2^{\text{OLS}} = \beta_2 + \sum a_t u_t$.
    - In order to $\hat{\beta}_2$ to be unbiased, $\operatorname{E}(\hat{\beta}_2) = \operatorname{E}\big(\beta_2 + \sum a_t u_t\big) \quad \Rightarrow \quad \operatorname{E}\big(\sum a_t u_t\big) = 0$.
        - If $a_t \perp u_t$, then $\operatorname{E}\big(\sum a_t u_t\big) = \sum\big[\operatorname{E}(a_t)\operatorname{E}(u_t)\big]$.
        - Given that $\operatorname{E}(u_t) = 0$, then $\operatorname{E}\big(\sum a_t u_t\big) = 0$, as long as a_t and u_t are uncorrelated.
        - Since a_t depends upon on all x_t (due to its denominator), the lack of correlation between a_t and u_t means a lack of correlation between the RHS variables and the error terms.
- However, y_t depends on y_{t-1}, which in turn is correlated with u_{t-1}: $\boldsymbol{y_{t-1}} = \beta_1 + \beta_2 y_{t-2} + \boldsymbol{u_{t-1}}$.
    - This means that there is correlation between a RHS variable ($x_{2t} = y_{t-1}$) and an error term (u_{t-1}).
    - This is not a correlation between contemporaneous RHS terms, though.
        - $x_{2t} = y_{t-1}$ and u_t are not correlated.

- This results in the presence of bias and might incur in loss of efficiency.
    - The estimator is still consistent, however.
- Note that in this case $x_{2t} = y_{t-1}$ is a random variable.
    - When the randomness of x_{2t} is correlated with the randomness of u_t, a problem arises.
    - Given that in this model $x_{2t} = y_{t-1}$, we know that correlation exists.

For the model:

$$y_t = \beta_1 + \beta_2 x_{2t} + u_t, \qquad x_{2t} = y_{t-1},\ u_t = \rho u_{t-1} + e_t$$

- There is correlation among the errors, for $\rho \neq 0$.
- There is also a contemporaneous correlation between the RHS variables and the error terms.
    - y_{t-1} is correlated with u_{t-1}, as is u_t.
    - Then there is correlation between non-contemporaneous ($x_{2t} = y_{t-1}$ and u_{t-1}) *and* contemporaneous terms ($x_{2t} = y_{t-1}$ and u_t).
- In this case, there is bias, inconsistency and inefficiency.
    - Consequently, the estimated residuals $\hat{u}_t$ will be biased.
    - The statistic for the Durbin-Watson test $d = \dfrac{\sum_{t=2}^{T} (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{T} \hat{u}_t^2}$ will thus be biased as well.
        - Durbin's h statistic (further explained next) should be used instead.

Durbin h statistic

The statistic is calculated as follows:

$$h = \hat{\rho} \sqrt{\frac{T}{1 - T \cdot \widehat{\mathrm{Var}}(\hat{\beta})}}$$

- $\hat{\rho} = 1 - \frac{1}{2} d$, where d is the statistic for the Durbin-Watson test.
- $\widehat{\mathrm{Var}}(\hat{\beta})$ is the variance of the parameter $\hat{\beta}$ associated with y_{t-1}.
- The statistic follows a standard normal distribution: $h \sim \mathcal{N}(0, 1)$.
- Note that this test will only work if $T \cdot \widehat{\mathrm{Var}}(\hat{\beta}) < 1$.
    - Otherwise, $\sqrt{\dfrac{T}{1 - T\widehat{\mathrm{Var}}(\hat{\beta})}} \notin \mathbb{R} \Rightarrow h \notin \mathbb{R}$.

Suppose a time series in which there is seasonality:

- For instance, the data for each month are very similar, throughout the years.
    - That would translate to $u_t = \rho u_{t-4} + e_t$, for quarterly data.
- Being a $\mathrm{AR}(1)$, the Durbin-Watson test would not be able to capture that behavior.
    - A more general test is needed.

Breusch-Godfrey LM Test

Consider the model:

$$
\begin{aligned}
y_t &= \beta_1 + \beta_2 x_{2t} + \ldots + \beta_k x_{kt} + u_t \\
u_t &= \rho_1 u_{t-1} + \ldots + \rho_p u_{t-p} + e_t
\end{aligned}
$$

Hypotheses:

$$
\begin{aligned}
&H_0 : \rho_1 = \rho_2 = \ldots = \rho_p = 0 \\
&H_1 : \exists\, \rho_j \neq 0,\ j = 1, 2, \ldots, p
\end{aligned}
$$

Steps:

1. Estimate y_t by OLS, save $\hat{u}_t$.
2. Regress $\hat{u}_t$ on constant, $x_{2t}, \ldots x_{kt}$ *and* $\hat{u}_{t-1}, \ldots, \hat{u}_{t-p}$.
    - The number of observations will be $T - p$: we will need to start from the $(p+1)^{\text{th}}$ observation.
3. Compute $(T - p) \cdot R^2$ (number of observations times the coefficient of determination), that will follow a χ^2 distribution with p restrictions (degrees of freedom).

More on the Breusch-Godfrey test here.

Correcting autocorrelation (quasi differencing):

For the model:

$$y_t = \beta_1 + \beta_2 x_{2t} + u_t, \quad u_t = \rho u_{t-1} + e_t$$

Lag the model one step back and multiply both sides by ρ :

$$\rho y_{t-1} = \rho \beta_1 + \rho \beta_2 x_{2,t-1} + \rho u_{t-1}$$

Now subtract the result from the original model:

$$y_t - \rho y_{t-1} = \beta_1 - \rho \beta_1 + \beta_2 x_{2t} - \rho \beta_2 x_{2,t-1} + u_t - \rho u_{t-1}$$

Since $u_t = \rho u_{t-1} + e_t$, it can be rewritten as:

$$y_t = \rho y_{t-1} + \beta_1 (1 - \rho) + \beta_2 x_{2t} - \rho \beta_2 x_{2,t-1} + e_t$$

- Now, there is not serial correlation among the errors any more.
- On the other hand, due to the interaction of ρ with the other terms, the model is no longer linear in the parameters.
    - This violates one of the Gauss-Markov assumptions (see Lecture 01).
    - Then, OLS should not be used to estimate the parameters.
        - Non-linear Least Squares (NLLS) should be used instead to estimate $\hat{\rho}$, $\hat{\beta}_1$ e $\hat{\beta}_2$.
- Were ρ known, the problem could be solved transforming the variables:
    - The model $(y_t - \rho y_{t-1}) = \beta_1 (1 - \rho) + \beta_2 (x_{2t} - \rho x_{2,t-1}) + e_t$ could be rewritten as $y_t^* = \beta_1 (1 - \rho) + \beta_2 x_{2t}^* + e_t$.
        - $\beta_1 (1 - \rho)$ can also be rewritten as β_1^*.
    - Without the non-linearity that stems from not knowing ρ (and consequently having to estimate it), the transformed model can be estimated using OLS.

- The transformed model has a different intercept ($\beta_1(1-\rho)$ rather than β_1), but the same slope (β_2).
- Due to the absence of serial correlation in the transformed model, the errors e_t are random - whereas the residuals u_t were not.

The same is applicable for higher-order autoregressive models:

- For instance, let the residuals be $u_t = \sum_{j=1}^{p} (\rho_j \cdot u_{t-j}) + e_t$.
- Then, for each ρ_j, we will lag the original model one step further back and subtract all the resultant p lagged models from the original.
    - The result will look like this: $y_t = \sum_{j=1}^{p} (\rho_j \cdot y_{t-j}) + \beta_1 \left(1 - \sum_{j=1}^{p} \rho_j\right) + \left[\beta_2 - \sum_{j=1}^{p} (\rho_j \cdot x_{2,t-j})\right] + \left[u_t - \sum_{j=1}^{p} (\rho_j \cdot u_{t-j})\right]$.
    - This could also be rewritten as $y_t = \sum_{j=1}^{p} (\rho_j \cdot y_{t-j}) + \ldots + \boldsymbol{e_t}$.