EC 508: Econometrics

Final Study Guide

Alex Hoagland, Boston University

May 2, 2019

Chapters 5 through 7

10. Dummy variables

$$y = \beta_0 + \boldsymbol{x}\beta + \gamma\mathbb{1}\{\text{something happens}\} + u$$

- γ measures the effect of a binary event
- Can be many binary events, as long as:
- **Dummy variable trap**: always need an omitted group, or else you have a problem of **perfect multicolinearity**

2. Adding variables to a model

Suppose that we are considering adding **one more variable** to a linear model—when will the **MSE decrease?**

1. Compare MSE_r with MSE_{ur}:

$$\frac{SSR_{ur}}{n-k-2} = MSE_{ur}, \quad \frac{SSR_r}{n-k-1} = MSE_r$$

2. Hence, the MSE decreases iff

$$\begin{aligned} MSE_{ur} \leq MSE_r &\Leftrightarrow (n-k-2+1)MSE_{ur} \leq (n-k-1)MSE_r \\ &\Leftrightarrow MSE_{ur} \leq (n-k-1)MSE_{ur} - (n-k-2)MSE_r \\ &\Leftrightarrow 1 \leq F \end{aligned}$$

3. Since we are testing inclusion of **one** new variable, $F = t^2$; hence, MSE goes down if $t^2 \geq 1$, or if $|t| \geq 1$

Other important things

- Be able to calculate how OLS estimates change when you scale data (Section 6.1)
- Know how to find optimal points for models with quadratic terms (p.194 in text)
- Understand how to interpret interaction effects (p.198)
- Understand difference between R^2 and $\overline{R}^2$ (p.202)
- Understand interpretation of linear probability models (Section 7.5)

Chapter 8: Heteroskedasticity

1. Heteroskedasticity

$$\mathbb{V}[u|x] \neq \sigma^2$$

Consequences:

- $\hat{\beta}$ are *still* unbiased and consistent
- $\overline{R}^2$ is unchanged
- Variances $\mathbb{V}[\hat{\beta}]$ are now biased $\rightarrow$ inference breaks down
- OLS is no longer BLUE

Under heteroskedasticity, $\mathbb{V}[u_i|x_i] = \sigma_i^2$

Recall that we can write the regression estimator as

$$\hat{\beta} = \beta + \frac{\sum_i (x_i - \overline{x})^2 u_i}{\text{SST}_x^2}$$

$$\Rightarrow \mathbb{V}[\hat{\beta}] = \frac{\sum_i (x_i - \overline{x})^2 \hat{u}_i}{\text{SST}_x^2}$$

This transformation of the residuals = **robust standard errors**

3. Breusch-Pagan test

We want to test the null hypothesis of **homoskedasticity**: $\mathcal{H}_0 : \mathbb{V}[u|x] = \sigma^2$

We want to test the null hypothesis of **homoskedasticity**: $\mathcal{H}_0 : \mathbb{V}[u|x] = \sigma^2$

We can rewrite this hypothesis as:

$$\begin{aligned}
\mathcal{H}_0 : \mathbb{V}[u|x] = \sigma^2 &\Leftrightarrow \mathbb{E}[u^2|x] = 0 \\
&\Leftrightarrow \mathbb{E}[u^2|x_1, x_2, ...] = 0 \\
&\Leftrightarrow u^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + ... + \delta_k x_k + \epsilon \\
&\Leftrightarrow \mathcal{H}_0' : \delta_1 = ... = \delta_k = 0. \qquad (*)
\end{aligned}$$

3. Breusch-Pagan test

We want to test the null hypothesis of **homoskedasticity**: $\mathcal{H}_0 : \mathbb{V}[u|x] = \sigma^2$

We can rewrite this hypothesis as:

$$\begin{aligned}
\mathcal{H}_0 : \mathbb{V}[u|x] = \sigma^2 &\Leftrightarrow \mathbb{E}[u^2|x] = 0 \\
&\Leftrightarrow \mathbb{E}[u^2|x_1, x_2, ...] = 0 \\
&\Leftrightarrow u^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + ... + \delta_k x_k + \epsilon \\
&\Leftrightarrow \mathcal{H}_0' : \delta_1 = ... = \delta_k = 0. \qquad (*)
\end{aligned}$$

When performed as an LM test (not an F-test), this is the **Breusch-Pagan** test

Test statistic: $LM = nR^2_* \sim \chi^2_k$

4. White test

Gauss-Markov still holds as long as $\text{Corr}(u^2, x_i) = \text{Corr}(u^2, x_i x_j) = 0$ for all i, j.

Hence, you can test **specifically** for heteroskedasticity that would invalidate OLS results using

$$\hat{u}^2 = \delta_0 + \left(\delta_1^1 x_1 + ... + \delta_k^1 x_k\right) + \left(\delta_1^2 x_1^2 + ... + \delta_k^2 x_k^2\right) + \left(\delta^3 x_1 x_2 + ... + \delta^3_{\binom{x}{2}} x_j x_{j'}\right) + \epsilon,$$

testing just as in the BP setting.

4. White test

Gauss-Markov still holds as long as $\text{Corr}(u^2, x_i) = \text{Corr}(u^2, x_i x_j) = 0$ for all i, j.

Hence, you can test **specifically** for heteroskedasticity that would invalidate OLS results using

$$\hat{u}^2 = \delta_0 + \left(\delta_1^1 x_1 + \ldots + \delta_k^1 x_k\right) + \left(\delta_1^2 x_1^2 + \ldots + \delta_k^2 x_k^2\right) + \left(\delta^3 x_1 x_2 + \ldots + \delta^3_{\binom{x}{2}} x_j x_{j'}\right) + \epsilon,$$

testing just as in the BP setting.

If you use the fitted values $\hat{y}$, this becomes $\hat{u}^2 = \delta_0 + \delta_1 \hat{y} + \delta_2 \hat{y}^2 + \epsilon$ —hence, can get **added power** for the test.

4. White test

Gauss-Markov still holds as long as $\mathrm{Corr}(u^2, x_i) = \mathrm{Corr}(u^2, x_i x_j) = 0$ for all i, j.

Hence, you can test **specifically** for heteroskedasticity that would invalidate OLS results using

$$\hat{u}^2 = \delta_0 + \left(\delta_1^1 x_1 + ... + \delta_k^1 x_k\right) + \left(\delta_1^2 x_1^2 + ... + \delta_k^2 x_k^2\right) + \left(\delta^3 x_1 x_2 + ... + \delta^3_{\binom{x}{2}} x_j x_{j'}\right) + \epsilon,$$

testing just as in the BP setting.

If you use the fitted values $\hat{y}$, this becomes $\hat{u}^2 = \delta_0 + \delta_1 \hat{y} + \delta_2 \hat{y}^2 + \epsilon$ —hence, can get **added power** for the test.

Warning! These tests require all other Gauss-Markov assumptions to hold to be meaningful (have to correctly specify the model)

What is WLS?

- Used before robust standard errors developed
- Requires knowing (or estimating) the form of heteroskedasticity: $\sigma^2 h(x) > 0$
- Minimizes the *weighted* sum of squared residuals

What is WLS?

- Used before robust standard errors developed
- Requires knowing (or estimating) the form of heteroskedasticity: $\sigma^2 h(x) > 0$
- Minimizes the *weighted* sum of squared residuals

If you know the weights:

$$\mathbb{E}\left(\frac{u_i}{\sqrt{h_i}}\right)^2 = \frac{\mathbb{E}(u_i^2)}{h_i} = \sigma^2$$

- This corrects SE-bias from heteroskedasticity
- Need only transform your variables by $\sqrt{h}^{-1}$ and run OLS

What if you need to estimate h?

Can estimate the form of heteroskedasticity using a (pretty) flexible form:

$$\mathbb{V}[u|x] = \sigma^2 \exp\{\delta_0 + \delta_1 x_1 + ... + \delta_k x_k\} \tag{1}$$

(This form ensures predicted variances are positive while maintaining generality)

What if you need to estimate h?

Can estimate the form of heteroskedasticity using a (pretty) flexible form:

$$\mathbb{V}[u|x] = \sigma^2 \exp\left\{\delta_0 + \delta_1 x_1 + ... + \delta_k x_k\right\} \tag{1}$$

(This form ensures predicted variances are positive while maintaining generality)

Estimation:

1. Rewrite (1) for estimating: $u^2 = \sigma^2 \exp\left\{\delta_0 + \delta_1 x_1 + ... + \delta_k x_k\right\} \nu$
2. Take logs: $\log(u) = a_0 + \delta_1 x_1 + ... + \delta_k x_k + \epsilon$
3. Run by OLS, obtain fitted values $\hat{g}$
4. Estimated weights are then $\hat{h}_i = \exp(\hat{g}_i)$

Chapter 10: Time Series Models

Static Models

$$y_t = \beta_0 + \beta_1 x_t + u_t, \; t \in \{1, 2, ..., n\}$$

- Contemporaneous effects only

Static Models

$$y_t = \beta_0 + \beta_1 x_t + u_t,\ t \in \{1, 2, ..., n\}$$

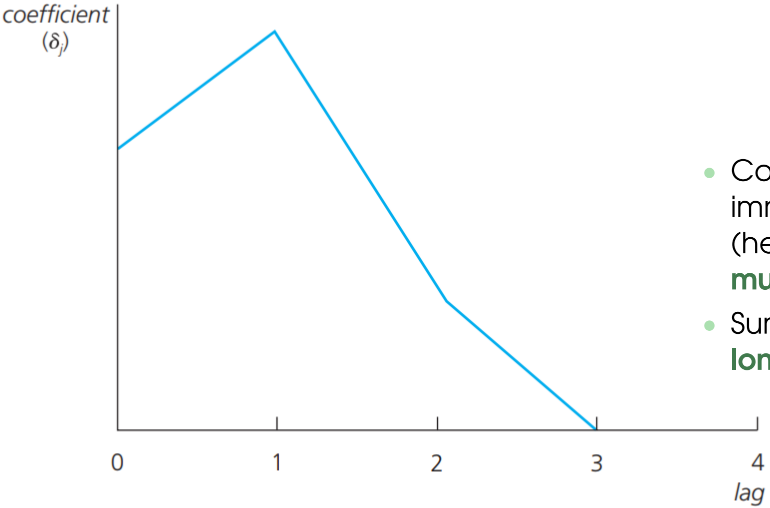
- Contemporaneous effects only

Finite distributed lag (FDL) models

$$y_t = \boldsymbol{x}_t \beta + \boldsymbol{x}_{t-j} \delta + u_t$$

- Lagged effects on y_t
- Nests static models if all δ coefficients are 0

8. Short- & long-run multipliers



- Coefficients give immediate impacts (hence δ_0 is **impact multiplier**)
- Sum of coefficients $\Rightarrow$ **long-run** impact (LRP)

9. Finite sample properties

How does **Gauss-Markov** work for time series?

- A1 (linear model) and A2 (no colinearity) don't change
- A3 (**strict** exogeneity) $\mathbb{E}[u_t|x_{t1},...,x_{tk}] = 0$ becomes $\mathbb{E}[u_t|x_{sj}] = 0$ for all s, j
  - This is a strong assumption (crime and police presence example)

How does **Gauss-Markov** work for time series?

- A1 (linear model) and A2 (no colinearity) don't change
- A3 (**strict** exogeneity) $\mathbb{E}[u_t|x_{t1},...,x_{tk}] = 0$ becomes $\mathbb{E}[u_t|x_{sj}] = 0$ for all s, j
  - This is a strong assumption (crime and police presence example)

These three assumptions $\Rightarrow$ OLS is **unbiased** (proof similar to original)

How does **Gauss-Markov** work for time series?

- A1 (linear model) and A2 (no colinearity) don't change
- A3 (**strict** exogeneity) $\mathbb{E}[u_t|x_{t1}, ..., x_{tk}] = 0$ becomes $\mathbb{E}[u_t|x_{sj}] = 0$ for all s, j
  - This is a strong assumption (crime and police presence example)

These three assumptions $\Rightarrow$ OLS is **unbiased** (proof similar to original)

- A4 (homoskedasticity) doesn't change
- A5 (no serial correlation—**new assumption**): $\text{Corr}(u_s, u_t) = 0$ for all $s \neq t$
  - Similar to the random sampling assumption

How does **Gauss-Markov** work for time series?

- A1 (linear model) and A2 (no colinearity) don't change
- A3 (**strict** exogeneity) $\mathbb{E}[u_t|x_{t1},...,x_{tk}] = 0$ becomes $\mathbb{E}[u_t|x_{sj}] = 0$ for all s, j
  - This is a strong assumption (crime and police presence example)

These three assumptions $\Rightarrow$ OLS is **unbiased** (proof similar to original)

- A4 (homoskedasticity) doesn't change
- A5 (no serial correlation—**new assumption**): $\text{Corr}(u_s, u_t) = 0$ for all $s \neq t$
  - Similar to the random sampling assumption

Under A1—A5, OLS is BLUE and the SEs are as previously calculated

10., 11., & 12. Dummies and logarithms

If you have all of the Gauss-Markov assumptions, **all** of the things you learned in cross-sectional analyses **carry over** to time series

If you have all of the Gauss-Markov assumptions, **all** of the things you learned in cross-sectional analyses **carry over** to time series

$$\log(y)_t = \beta_0 + \beta_1 \mathbb{1}\{z_t > 0\} + \delta_0 x_t + \delta_1 x_{t-1} + ... + \delta_k x_{t-k} + u_t$$

If you have all of the Gauss-Markov assumptions, **all** of the things you learned in cross-sectional analyses **carry over** to time series

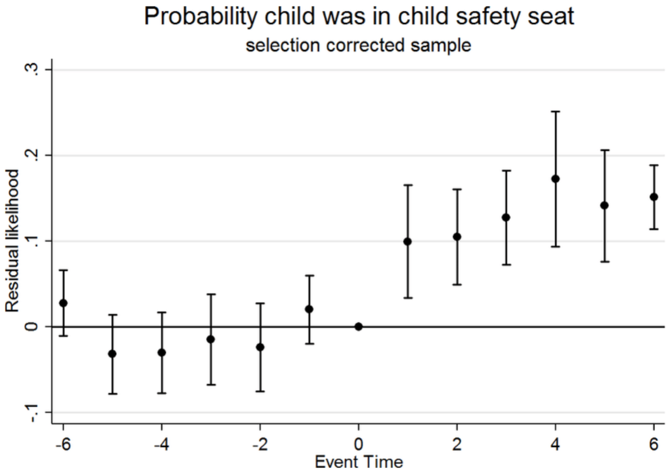
$$\log(y)_t = \beta_0 + \beta_1 \mathbb{1}\{z_t > 0\} + \delta_0 x_t + \delta_1 x_{t-1} + \ldots + \delta_k x_{t-k} + u_t$$

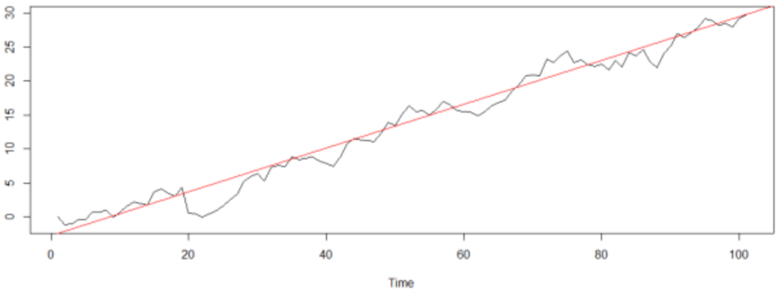
- β_1 estimates the effect of a certain outcome in a given period
  - Used to perform event study analysis (how does outcome evolve?)

If you have all of the Gauss-Markov assumptions, **all** of the things you learned in cross-sectional analyses **carry over** to time series

$$\log(y)_t = \beta_0 + \beta_1 \mathbb{1}\{z_t > 0\} + \delta_0 x_t + \delta_1 x_{t-1} + ... + \delta_k x_{t-k} + u_t$$

- β_1 estimates the effect of a certain outcome in a given period
  - Used to perform event study analysis (how does outcome evolve?)
- δ_0 is the **short-run elasticity** of y with respect to x
- $\delta_0 + \delta_1 + ... + \delta_k$ is the **long-run elasticity**


Probability child was in child safety seat
selection corrected sample
Residual likelihood
-.1
0
.1
.2
.3
-6
-4
-2
0
2
4
6
Event Time



- Ignoring time trends $\Rightarrow$ spurious correlations
- Want to detrend if any variable is trending
- Control for this by including t as a regressor, or by an **exponential** trend of the form $\log(y)_t = \beta_0 + \boldsymbol{x}_t\beta + \alpha t + u_t$ ($\beta_1 \approx$ growth rate)

14. Seasonality

- Control for seasonality with season dummies—**careful of the dummy variable trap!**

Chapter 11: More on Time Series

15. Stationarity

To analyze dynamic models, we need an idea of **how** the random process evolves over time. The simplest (and strongest) requirement:

The joint dist. of $(x_{t_1}, x_{t_2}, ..., x_{t_m})$ is the same as the dist. of $(x_{t_1+h}, x_{t_2+h}, ..., x_{t_m+h})$

- **All** of a distribution's moments are constant over time
- Usually **very restrictive**—can limit to **two** moments if distribution has a finite second moment

15. Stationarity

To analyze dynamic models, we need an idea of **how** the random process evolves over time. The simplest (and strongest) requirement:

The joint dist. of $(x_{t_1}, x_{t_2}, ..., x_{t_m})$ is the same as the dist. of $(x_{t_1+h}, x_{t_2+h}, ..., x_{t_m+h})$

- **All** of a distribution's moments are constant over time
- Usually **very restrictive**—can limit to **two** moments if distribution has a finite second moment

Covariance/weak stationarity:

1. $\mathbb{E}[x_t]$ is constant
2. $\mathbb{V}[x_t]$ is constant
3. For all t and $h \geq 1$, $\text{Cov}(x_t, x_{t+h}) = g(h)$ does not depend on t

16. Weakly dependent time-series

Weak dependence controls how much the distant past controls the present

- Different from stationarity, which discusses **joint distributions** of outcomes

Weak Dependence in time series is approximately:

$$\lim_{h_t \to \infty} \text{Corr}(x_t, x_{t+h}) = 0$$

- Series with this property are referred to as **asymptotically uncorrelated**
- Both stationarity and dependence are important for **inference**

$$x_t = e_t + \alpha_1 e_{t-1} + \alpha_2 e_{t-2} + ... + \alpha_q e_{t-q}$$

What makes this weakly dependent?

- $\text{Cov}(x_t, x_{t+1}) = \text{Cov}(e_t + \alpha_1 e_{t-1}, e_{t+1} + \alpha_1 e_t) = \alpha_1 \sigma_e^2$
- But if you go father, $\text{Cov}(x_t, x_{t+j}) = \text{Cov}(e_t + \alpha_1 e_{t-1}, e_{t+j} + \alpha_1 e_{t+j-1}$ = 0
- Hence, this process is weakly dependent (and stationary)

$$y_t = \rho_1 y_{t-1} + \rho_2 y_{t-2} + ... + \rho_p y_{t-p} + e_t$$

What makes this weakly dependent?

- Only work with **stable processes**, so $|\rho_1| < 1$
- Also assume AR(1) is covariance stationary (complicated proof)
- Then $\mathbb{V}[y_t] = \rho_1^2 \mathbb{V}[y_{t-1}] + \sigma_e^2$, and since variance constant: $\sigma_y^2 = \rho_1^2 \sigma_y^2 + \sigma_e^2$
- Next, solve y_{t+h} backwards: $y_{t+h} = \rho_1^h y_t + \rho_1^{h-1} e_{t+1} + ... + \rho_1 e_{t+h-1} + e_{t+h}$
- Multiply both sides by y_t and take expectations:

$$\begin{aligned} \text{Cov}(y_t, y_{t+h}) &= \mathbb{E}\left[\rho_1^h y_t^2 + \rho_1^{h-1} e_{t+1} y_t + ... + \rho_1 e_{t+h-1} y_t + e_{t+h} y_t\right] \\ &= \rho_1^h \sigma_y^2 \end{aligned}$$

$$y_t = \rho_1 y_{t-1} + \rho_2 y_{t-2} + \ldots + \rho_p y_{t-p} + e_t$$

What makes this weakly dependent?

- Only work with **stable processes**, so $|\rho_1| < 1$
- Also assume AR(1) is covariance stationary (complicated proof)
- Then $\mathbb{V}[y_t] = \rho_1^2 \mathbb{V}[y_{t-1}] + \sigma_e^2$, and since variance constant: $\sigma_y^2 = \rho_1^2 \sigma_y^2 + \sigma_e^2$
- Next, solve y_{t+h} backwards: $y_{t+h} = \rho_1^h y_t + \rho_1^{h-1} e_{t+1} + \ldots + \rho_1 e_{t+h-1} + e_{t+h}$
- Multiply both sides by y_t and take expectations:

$$\begin{aligned} \text{Cov}(y_t, y_{t+h}) &= \mathbb{E}\left[\rho_1^h y_t^2 + \rho_1^{h-1} e_{t+1} y_t + \ldots + \rho_1 e_{t+h-1} y_t + e_{t+h} y_t\right] \\ &= \rho_1^h \sigma_y^2 \end{aligned}$$

Since $|\rho_1| < 1$, $\rho_1^h \rightarrow_{h\rightarrow\infty} 0$!

19. & 20. Highly persistent series: random walks

If time series are **weakly dependent**, then Gauss Markov assumptions hold. But what if they aren't? What if effects are long-lasting?

A **random walk** is an AR(1) with $\rho_1 = 1$:

$$y_t = y_{t-1} + e_t$$

- Can show that $\mathbb{E}[y_t] = \mathbb{E}[y_0]$ for all t, but that $\mathbb{V}[y_t] = \sigma_e^2 t$
- Work on previous slide shows this isn't weakly dependent either!
- Easy fix—"integrate" out persistence using first differences

$$\Delta y_t = (y_{t-1} + e_t) - (y_{t-1}) = e_t$$

21. Random walks with drift & stationarity

$$y_t = \alpha_0 + y_{t-1} + e_t$$

- Now, $\mathbb{E}[y_t] = \alpha_0 t + \mathbb{E}[y_0]$, so not stationary!
- Again, differencing out linearly works:

$$\Delta y_t = (\alpha_0 + \alpha_1 y_{t-1} + e_t) - y_{t-1} = \alpha_0 + e_t$$

21. Random walks with drift & stationarity

$$y_t = \alpha_0 + y_{t-1} + e_t$$

- Now, $\mathbb{E}[y_t] = \alpha_0 t + \mathbb{E}[y_0]$, so not stationary!
- Again, differencing out linearly works:

$$\Delta y_t = (\alpha_0 + \alpha_1 y_{t-1} + e_t) - y_{t-1} = \alpha_0 + e_t$$

As a rule, **first differencing** is handy. If linear differencing isn't sufficient, **log differencing** usually is:

$$\Delta \log(y_t) = \log(y_t) - \log(y_{t-1})$$

Most highly persistent time series are examples of **unit root proceses**: a mix of an AR(1) term with $\rho = 1$ and any **weakly dependent** process

How do you know if a time series is I(1)?

1. Test the first-order autocorrelation—if $\rho_1 \approx 1$, you may have a problem (tests are hard here)
2. Dickey-Fuller test: fit $y_t = \alpha_0 + \rho y_{t-1} + \delta t$ and test $\mathcal{H}_0 : \rho = 1$

```
. dfuller ln_divyield

Dickey-Fuller test for unit root                   Number of obs   =       863

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -2.054            -3.430            -2.860            -2.570
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.2635
```

Chapter 12: Serial Correlation

23. What happens under serial correlation?

Usually, we are working with a **dynamic model** of the form

$$y_t = \boldsymbol{x}_t\beta + \boldsymbol{y}_{t-h}\delta + u.$$

23. What happens under serial correlation?

Usually, we are working with a **dynamic model** of the form

$$y_t = \boldsymbol{x}_t\beta + \boldsymbol{y}_{t-h}\delta + u.$$

Serial correlation occurs when $\text{Cov}(u_t, u_{t+h}) \neq 0$

- This **always** messes up your standard errors (similar to heteroskedasticity)
- Also, this **might** lead to inconsistent parameter estimates if your dynamic model is not correctly specified

24. Detecting serial correlation

In general, simplest test is that u_t follows an **AR(1) process**: $\mathcal{H}_0 : \rho_u = 0$.

When all regressors are strictly exogenous:

1. Estimate dynamic model and obtain $\{\hat{u}_t\}$
2. Estimate $\hat{u}_t = \rho \hat{u}_{t-1} + e_t$
   - Note that there is **no intercept here**
3. Perform a t test for $\mathcal{H}_0$

In general, simplest test is that u_t follows an **AR(1) process**: $\mathcal{H}_0 : \rho_u = 0$.

When all regressors are strictly exogenous:

1. Estimate dynamic model and obtain $\{\hat{u}_t\}$
2. Estimate $\hat{u}_t = \rho \hat{u}_{t-1} + e_t$
   - Note that there is **no intercept here**
3. Perform a t test for $\mathcal{H}_0$

If regressors aren't strictly exogenous:

1. Same as (1) above
2. Estimate $\hat{u}_t = \boldsymbol{x}_t \beta + \rho \hat{u}_{t-1}$
3. Perform another t test on $\hat{\rho}$

25. Correcting for serial correlation

We can use **feasible GLS** to correct for serial correlation if ρ is known (or estimated):

1. Estimate $y_t = \boldsymbol{x}_t\beta$ and obtain residuals
2. Estimate residual equation (step 2 in one of the methods above) to obtain $\hat{\rho}$
3. **Quasi-difference** your data: $\tilde{z}_t = z_t - \hat{\rho}z_{t-1}$ for all variables
4. Use OLS to estimate $\tilde{y}_t = \tilde{\boldsymbol{x}}_t\beta + u$

In Stata, these are called **Cochrane-Orcutt** or **Prais-Winston** estimators

Questions?