Introductory Econometrics I

Intra-Cluster and Serial Correlation

Yingjie Feng

School of Economics and Management

Tsinghua University

May 10, 2024

Outline

Review: Assumptions for OLS

- Recall the classical linear model (CLM) assumptions for OLS regression:
  - MLR.1: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + u$
  - MLR.2: **random sampling** from the population
  - MLR.3: no perfect collinearity in the sample
  - MLR.4: $\mathbb{E}[u|x_1, ..., x_k] = \mathbb{E}[u] = 0$ (exogenous explanatory variables)
  - MLR.5: $\mathbb{V}[u|x_1, ..., x_k] = \mathbb{V}[u] = \sigma^2$ (homoskedasticity)
  - MLR.6: $u|x_1, \cdots, x_k \sim \mathsf{Normal}(0, \sigma^2)$
- We know MLR.6 is *not* necessary for large sample inference
- We also have relaxed MLR.5 and allow for *heteroskedasticity*-robust inference
- MLR.4 is a key identifying condition (we will relax it)
- MLR.3 is a mild condition for the existence of OLS estimators

Relaxing the Independence Assumption

- Let us focus on MLR.2: data are *independent* and identically distributed
  - $(y_i, x_{i1}, \cdots, x_{ik})$ or $(x_{i1}, \cdots, x_{ik}, u_i)$ are independent over i
- However, in many scenarios this is unrealistic. For example,
  - Intra-cluster correlation: data with "group" structure (Chapter 14)
    - Cities within the same province are correlated
    - The same person observed in several periods is correlated (panel data)
    - Trade flows between countries (network data, more complex)
  - Serial correlation: time series data (Chapter 12)
    - Observations in different periods may be dependent (e.g. GDP growth in 2021 may depends on GDP growth in 2020)
    - Similar (not the same) logic may also apply to spatial econometric analysis

Relaxing the Independence Assumption

- What happens if data are dependent?
- Generally, under certain conditions, we still have the unbiasedness, consistency and asymptotic normality of OLS estimators
  - We did not relax Assumption MLR.4 (though we need a stronger version)
  - But we need to guarantee we are using the "correct" variance formula (or standard errors)
  - Similar to heteroskedasticity-robust inference, we may want to do inference robust to intra-cluster/serial correlation
- Again, we can also construct other estimators (e.g., generalized least squares) to improve efficiency (not discussed in this class).

Revisit: Variance for Simple Regression

- To see the problem at hand, consider the simple regression model

$$y_i = \beta_0 + \beta_1 x_i + u_i, \quad \mathbb{E}[u_i|x_i] = 0$$

  - Caveat: for dependent data, assuming $\mathbb{E}[u_i|x_i] = 0$ is often not enough for unbiasedness, but for the moment, we focus on variance calculation
  - OLS estimator $\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^n (x_i - \bar{x}) u_i}{\sum_{i=1}^n (x_i - \bar{x})^2}$

- Conditional on $\{x_i\}_{i=1}^n$,

$$\mathbb{V}[\hat{\beta}_1] = \mathbb{V}\left[\frac{\sum_{i=1}^n (x_i - \bar{x}) u_i}{\sum_{i=1}^n (x_i - \bar{x})^2}\right] = \frac{\mathbb{V}[\sum_{i=1}^n (x_i - \bar{x}) u_i]}{(\sum_{i=1}^n (x_i - \bar{x})^2)^2}$$

$$= \underbrace{\frac{\sum_{i=1}^n (x_i - \bar{x})^2 \mathbb{V}[u_i]}{(\sum_{i=1}^n (x_i - \bar{x})^2)^2}}_{\text{heteroskedasticity-robust}} + \underbrace{\frac{2\sum_{i=1}^{n-1} \sum_{j=i+1}^n \mathbb{C}ov[(x_i - \bar{x}) u_i, (x_j - \bar{x}) u_j]}{(\sum_{i=1}^n (x_i - \bar{x})^2)^2}}_{\neq 0 \text{ if data are dependent}}$$

Outline

Clustered Data

- Data that can be classified into a number of distinct groups or "clusters"
  - Emloyee-level data: they belong to different firms
  - County-level data: they belong to different prefectural cities
  - Panel data: individual-year observations belong to different individuals
- In general, we can write the data as

$$y_{i,g}, x_{i1,g}, \cdots, x_{ik,g}, \quad g = 1, \cdots, G, \quad i = 1, \cdots, n_g$$

  - g indexes groups; i indexes a unit within each group g
  - In total, we have G groups (n_g units in group g)
- Multiple regression model

$$y_{i,g} = \beta_0 + \beta_1 x_{i1,g} + \cdots + x_{ik,g} + u_{i,g}$$

- Independence of all observations may be unrealistic.
  - Units within the same cluster may share some common features/shocks

Intra-Cluster Correlation

- It makes more sense to assume

$$\mathbb{C}ov[u_{i,g}, u_{j,g}] \neq 0, \quad i \neq j$$

$$\mathbb{C}ov[u_{i,g}, u_{j,l}] = 0, \quad g \neq l$$

  - units within the same cluster are correlated
  - units in different clusters are uncorrelated

- Consider the simple OLS estimators $\hat{\beta}_1$ from regressing $y_{i,g}$ on $x_{i,g}$

$$\mathbb{V}[\hat{\beta}_1] = \underbrace{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 \mathbb{V}[u_i]}{(\sum_{i=1}^{n}(x_i - \bar{x})^2)^2}}_{\text{heteroskedasticity-robust}} + \underbrace{\frac{2\sum_{i=1}^{n-1}\sum_{j>i}\mathbb{C}ov[(x_i - \bar{x})u_i, (x_j - \bar{x})u_j]}{(\sum_{i=1}^{n}(x_i - \bar{x})^2)^2}}_{\neq 0 \text{ if data are dependent}}$$

$$= \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 \mathbb{V}[u_i]}{(\sum_{i=1}^{n}(x_i - \bar{x})^2)^2} + \frac{2\sum_{g=1}^{G}\sum_{\substack{i,j \in \mathcal{C}_g \\ i<j}} \mathbb{C}ov[(x_i - \bar{x})u_i, (x_j - \bar{x})u_j]}{(\sum_{i=1}^{n}(x_i - \bar{x})^2)^2}$$

$\mathcal{C}_g$ denotes the gth cluster

Intra-Cluster Correlation

- The variance formula for more general multiple regression is also available
- Like heteroskedasticity-robust inference, we need to replace $\mathbb{V}[u_i]$ and $\mathbb{C}ov[u_i, u_j]$ with something known
  - Replace $\mathbb{V}[u_i]$ with $\hat{u}_i^2$ and $\mathbb{C}ov[u_i, u_j]$ with $\hat{u}_i \hat{u}_j$
  - Again, they are not "good" estimates for each of these parameters
  - But the resulting estimator $\hat{\mathbb{V}}[\hat{\beta}_1]$ is a good estimate of $\mathbb{V}[\hat{\beta}_1]$ (sample averaging and LLN help)
- In Stata,
  - `reg y x1 x2, vce(cluster clusterid)`
- Reminder:
  - It is robust to both heteroskedasticity and intra-cluster correlation
  - It is robust to *unknown* intra-cluster correlation structure
  - A finite sample adjustment is used by Stata command

Intra-Cluster Correlation: Example

- Impact of "tracking" on educational attainment

$$score_{i,g} = \beta_0 + \beta_1 \; tracking_g + u_{i,g}$$

  - $score_{i,g}$: standardized score of student i in school g
  - $tracking_g = 1$ if school g assigns students based on initial test score; $tracking_g = 0$ otherwise
  - local demographics, individual teachers, etc. may affect students' performance, implying intra-cluster (within-school) correlation
  - But it may be reasonable to assume students achievement in different schools are independent

Intra-Cluster Correlation: Example

- Standardized test score and run simple regression with conventional s.e. (assuming homoskedasticity and uncorrelatedness)

```
. egen testscore = std(totalscore)

. reg testscore tracking

      Source |       SS           df       MS      Number of obs   =     5,795
-------------+----------------------------------   F(1, 5793)      =     27.73
       Model |  27.6035514         1  27.6035514   Prob > F        =    0.0000
    Residual |  5766.39646     5,793  .99540764   R-squared       =    0.0048
-------------+----------------------------------   Adj R-squared   =    0.0046
       Total |  5794.00001     5,794           1   Root MSE        =     .9977

------------------------------------------------------------------------------
   testscore |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    tracking |   .1380913   .0262231     5.27   0.000     .0866842    .1894983
       _cons |  -.0710354   .0188078    -3.78   0.000    -.1079057   -.034165
------------------------------------------------------------------------------
```

Intra-Cluster Correlation: Example

- s.e. robust to heteroskedasticity (but not intra-cluster correlation)

```
. reg testscore tracking, robust

Linear regression                               Number of obs     =      5,795
                                                F(1, 5793)        =      27.76
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0048
                                                Root MSE          =      .9977
```

testscore	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
tracking	.1380913	.0262102	5.27	0.000	.0867095	.189473
_cons	-.0710354	.0186418	-3.81	0.000	-.1075803	-.0344904

Intra-Cluster Correlation: Example

- s.e. robust to heteroskedasticity and intra-cluster correlation

```
. reg testscore tracking, vce(cluster schoolid)

Linear regression                               Number of obs     =      5,795
                                                F(1, 120)         =       3.20
                                                Prob > F          =     0.0763
                                                R-squared         =     0.0048
                                                Root MSE          =      .9977

                            (Std. Err. adjusted for 121 clusters in schoolid)
```

testscore	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
tracking	.1380913	.0772362	1.79	0.076	-.0148311	.2910136
_cons	-.0710354	.0543934	-1.31	0.194	-.1787304	.0366597

- The same as `reg y x, cluster(schoolid)`
- **Reminder**: In this example, clustering increases s.e. and decreases significance; but in general, cluster-robust s.e. could go either way.

Cluster-Level Latent Variables

- Now consider a particular scenario where clustering arises due to the cluster-level *latent variables.* For example,

$$y_{i,e} = \beta_0 + \beta_1 x_{i,e} + f_i + v_{i,e}, \quad u_{i,e} = f_i + v_{i,e}$$

  - i indexes firms; e indexes employees (in this case i denotes **clusters**)
  - f_i: unobserved firm-level characteristics (e.g., management level, risk preference, etc.)
  - We leave f_i in the error term $u_{i,e}$
  - $(x_{i,e}, v_{i,e})$ are i.i.d. over i and e; f_i are i.i.d. over i only
  - Employees in the same firm share the same f_i!
  - So $u_{i,e}$ are independent over i, but not over e

Cluster-Level Latent Variables

- Now consider a particular scenario where clustering arises due to the cluster-level *latent variables*. For example,

$$y_{i,e} = \beta_0 + \beta_1 x_{i,e} + f_i + v_{i,e}, \quad u_{i,e} = f_i + v_{i,e}$$

  - We may still have unbiasedness and consistency of OLS if we assume

$$\mathbb{E}[v_{i,e}|(x_{i,1}, \cdots, x_{i,n_i})] = 0, \quad \mathbb{E}[f_i|(x_{i,1}, \cdots, x_{i,n_i})] = 0$$

    - Recall that when we prove unbiasedness of $\hat{\beta}_1$, we actually need

$$\mathbb{E}[u_{i,e}|\boldsymbol{X}] = 0$$

    - $\boldsymbol{X}$ contains all explanatory variables $\{(x_{i,1}, \cdots, x_{i,n_i}) : 1 \leq i \leq G\}$
    - Independence across clusters means we can safely just assume

$$\mathbb{E}[u_{i,e}|x_{i,1}, \cdots, x_{i,n_i}] = 0$$

Cluster-Level Latent Variables

- Now consider a particular scenario where clustering arises due to the cluster-level *latent variables*. For example,

$$y_{i,e} = \beta_0 + \beta_1 x_{i,e} + f_i + v_{i,e}, \quad u_{i,e} = f_i + v_{i,e}$$

  - We may still have unbiasedness and consistency of OLS if we assume

$$\mathbb{E}[v_{i,e}|(x_{i,1}, \cdots, x_{i,n_i})] = 0, \quad \mathbb{E}[f_i|(x_{i,1}, \cdots, x_{i,n_i})] = 0$$

  - Omitting f_i does not cause bias since it is mean independent of the explanatory variables
  - In this case, f_i is usually known as *random effect* in panel data literature
  - By contrast, if f_i is correlated with explanatory variables, f_i is often called *fixed effects* (not covered in this class, but see Wooldridge Chapter 14)

Cluster-Level Latent Variables

- Now consider a particular scenario where clustering arises due to the cluster-level *latent variables*. For example,

$$y_{i,e} = \beta_0 + \beta_1 x_{i,e} + f_i + v_{i,e}, \quad u_{i,e} = f_i + v_{i,e}$$

  - To construct confidence intervals or conduct hypothesis testing, we need to calculate the correct variance
  - Assume $\mathbb{V}[f_i] = \sigma_f^2$, $\mathbb{V}[v_{i,e}] = \sigma_e^2$, $\mathbb{C}ov[f_i, v_{i,e}] = 0$

$$\mathbb{V}[u_{i,e}] = \mathbb{V}[f_i] + \mathbb{V}[v_{i,e}] = \sigma_f^2 + \sigma_v^2, \quad \mathbb{C}ov[u_{i,e}, u_{j,g}] = 0 \text{ for } i \neq j,$$

$$\begin{aligned}
\mathbb{C}ov[u_{i,e}, u_{i,g}] &= \mathbb{C}ov[f_i + v_{i,e}, f_i + v_{i,g}] \\
&= \mathbb{C}ov[f_i, f_i] + \mathbb{C}ov[v_{i,e}, v_{i,g}] + \mathbb{C}ov[v_{i,e}, f_i] + \mathbb{C}ov[f_i, v_{i,g}] \\
&= \sigma_f^2 \qquad \text{for } e \neq g.
\end{aligned}$$

  - Cluster-robust inference is more reasonable, but a more efficient GLS estimator is proposed in panel data literature (random-effect estimator)

Intra-Cluster Correlation: Final Remarks

- Allowing for arbitrary correlation within cluster indeed reduces the "effective sample size"
- The total number of observations $n = \sum_{g=1}^{G} n_g$, but you only have G independent clusters!
  - Theoretically, usual large sample asymptotics need $G \to \infty$ (many clusters)
  - Be aware of G in practice, especially when it is small!
- At what level to cluster? A bias-variance trade-off for variance estimators:
  - Cluster at lower-level (less aggregate): variance estimate may be biased, but with more clusters it is less variable
  - Cluster at higher-level (more aggregate): variance estimate may be unbiased, but with fewer clusters it could have high variability (low precision)

Outline

1. Introduction: Relaxing the Independence Assumption

2. Clustered Data and Intra-Cluster Correlation

3. Time Series Data and Serial Correlation
   - Standard Errors Robust to Serial Correlation
   - Examples of Weakly Dependent Time Series*

Outline

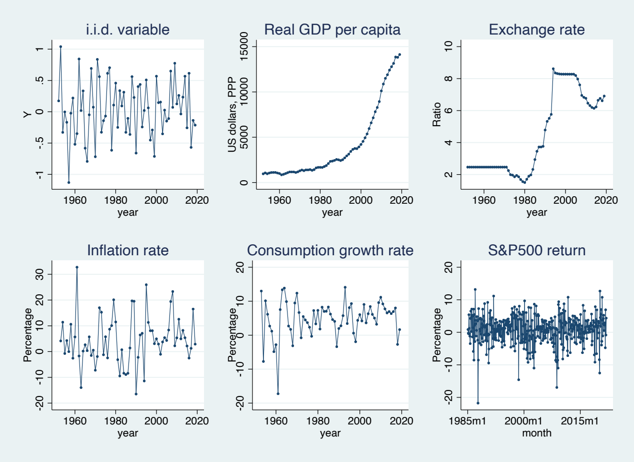
Time Series Data

- Differences between time series (TS) data and cross sectional (CS) data

  1. Time series data come with a temporal ordering (from earliest to latest). Usually, the ordering of the data is important.
  2. We cannot think of time series data as a random sample of units (individuals, firms, etc.) from a large population. So we cannot realistically impose random sampling (Assumption MLR.2).
  3. In fact, time series data are almost always correlated across time, sometimes very strongly.
     - ⋆ Very strong time series correlation (persistence) can cause problems for the usual ordinary least squares inference (not covered in this class).
     - ⋆ Many time series data exhibit *trends*; For time series at the monthly and quarterly frequencies, *seasonality* can be an issue. (not covered in this class)

Time Series Data

Some examples:



Time Series Regression

- Here we only consider "weakly dependent" data for which the usual OLS estimators have similar properties we had for cross sectional data
- Consider multiple regression

$$y_t = \beta_0 + \beta_1 x_{t1} + ... + \beta_k x_{tk} + u_t, \quad t = 1, \cdots, T$$

  - Sample size is the number of time periods T
- For the moment, we again *only* focus on the calculation of variance
- Formal and comprehensive treatment of time series analysis will be covered in Introductory Econometrics II

Serial Correlation

- Serial correlation means that the errors $\{u_t : t = 1, 2, ...\}$ are correlated.

- Like intra-cluster correlation, serial correlation has nothing directly to do with unbiasedness or consistency of OLS.

- But serially correlated errors means the usual OLS statistical inference is incorrect, even in large samples. In many cases, the inference can be very misleading.

Standard Errors Robust to Serial Correlation

- It is common to treat serial correlation in TS regression like we often treat heteroskedasticity in CS regression
  - Still use OLS estimators, but correct the inference procedure
  - But it is also possible to test for serial correlation and use other estimators, e.g., generalized least squares (not covered in class)
- Recall we can make inference robust to heteroskedasticity of unknown form
- It is also possible to compute standard errors, CIs, and test statistics robust to general forms of serial correlation – at least approximately (they are also robust to any kind of heteroskedasticity).
- The underlying theory is complicated, but the basic idea is the same as before (allow u_t to be correlated with some u_{t+j})

Standard Errors Robust to Serial Correlation

- In general, assume

$$\gamma_j = \mathbb{C}ov[u_t, u_{t+j}] \neq 0$$

  - Intuition of "weak dependence": Observations far away from each other in time are approximately uncorrelated ($\gamma_j \to 0$ as $j \to \infty$)

- Consider the simple OLS estimators $\hat{\beta}_1$ from regressing y_t on x_t ($k = 1$):

$$\mathbb{V}[\hat{\beta}_1] = \underbrace{\frac{\sum_{t=1}^{T}(x_t - \bar{x})^2 \mathbb{V}[u_t]}{(\sum_{t=1}^{T}(x_t - \bar{x})^2)^2}}_{\text{heteroskedasticity-robust}} + \underbrace{\frac{2\sum_{t=1}^{T-1}\sum_{j=1}^{T-t} \mathbb{C}ov[(x_t - \bar{x})u_t, (x_{t+j} - \bar{x})u_{t+j}]}{(\sum_{t=1}^{T}(x_t - \bar{x})^2)^2}}_{\neq 0 \text{ if data are dependent}}$$

$$= \frac{\sum_{t=1}^{T}(x_t - \bar{x})^2 \mathbb{V}[u_t]}{(\sum_{t=1}^{T}(x_t - \bar{x})^2)^2} + \frac{2\sum_{t=1}^{T-1}\sum_{j=1}^{T-t}(x_t - \bar{x})(x_{t+j} - \bar{x})\gamma_j}{(\sum_{t=1}^{T}(x_t - \bar{x})^2)^2}$$

$$\approx \frac{\sum_{t=1}^{T}(x_t - \bar{x})^2 \mathbb{V}[u_t]}{(\sum_{t=1}^{T}(x_t - \bar{x})^2)^2} + \frac{2\sum_{t=1}^{T-1}\sum_{j=1}^{M}(x_t - \bar{x})(x_{t+j} - \bar{x})\gamma_j}{(\sum_{t=1}^{T}(x_t - \bar{x})^2)^2}$$

Standard Errors Robust to Serial Correlation

- Since $\mathbb{E}[u_{t+j}u_t] \approx 0$ when j is large, truncate at distance $j = M$ is a valid approximation under some technical conditions
  - We do not account for correlation between $x_t u_t$ and $x_{t+j}u_{t+j}$ for $j > M$
- The variance formula for more general multiple regression is also available
- Like cluster-robust inference, we need to replace $\mathbb{V}[u_t]$ and $\mathbb{C}ov[u_t, u_{t+j}]$ with something known
- **Reminder**: This is only an imprecise description of the main idea. Some adjustments are needed for good performance

Standard Errors Robust to Serial Correlation

- The resulting standard errors are called **Newey-West standard errors**
- Such standard errors are sometimes called **HAC (heteroskedasticity and autocorrelation consistent) standard errors**.
- The N-W standard errors are not as automated as the adjustment for heteroskedasticity
  - We have to choose the lag M.
  - With annual data, the lag is usually fairly short (maybe a couple of years like $M = 2$); with quarterly or monthly data we try longer lags (like $M = 24$)
  - In practice, we probably experiment a bit to see how sensitive the standard errors are.

Standard Errors Robust to Serial Correlation

- In Stata,
  - `newey y x1 x2 ... xk, lag(M)`
  - Setting $M = 0$ is the same as
    - `reg y x1 x2 ... xk, robust`
- Important: We are still estimating the parameters by OLS estimators. We are only changing how we estimate their precision and perform inference.
- Just like heteroskedasticity-robust inference, we can apply the HAC inference whether or not we have evidence of serial correlation.

Standard Errors Robust to Serial Correlation

- PRMINWGE.DTA: Effect of (log) minimum wage (*lmincov*) on employment rate (*lprepop*)

```
. reg lprepop lmincov lusgnp lprgnp t, robust

Linear regression                               Number of obs     =         38
                                                F(4, 33)          =      55.83
                                                Prob > F          =     0.0000
                                                R-squared         =     0.8892
                                                Root MSE          =     .03277

------------------------------------------------------------------------------
             |               Robust
     lprepop |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     lmincov |  -.2122612      .04239    -5.01   0.000    -.2985043   -.1260182
      lusgnp |   .4860463    .2387616     2.04   0.050     .0002822    .9718105
      lprgnp |   .2852386     .089022     3.20   0.003     .1041219    .4663553
           t |  -.0266633    .0048954    -5.45   0.000     -.036623   -.0167036
       _cons |  -6.663432    1.303295    -5.11   0.000    -9.315005   -4.011859
------------------------------------------------------------------------------
```

Standard Errors Robust to Serial Correlation

- PRMINWGE.DTA: Effect of (log) minimum wage (*lmincov*) on employment rate (*lprepop*)

```
. tsset year
        time variable:  year, 1950 to 1987
                delta:  1 unit

. newey lprepop lmincov lusgnp lprgnp t, lag(0)

Regression with Newey-West standard errors      Number of obs     =         38
maximum lag: 0                                  F(  4,        33) =      55.83
                                                Prob > F          =     0.0000
```

lprepop	Coef.	Newey-West Std. Err.	t	P>|t|	[95% Conf.	Interval]
lmincov	-.2122612	.04239	-5.01	0.000	-.2985043	-.1260182
lusgnp	.4860463	.2387616	2.04	0.050	.0002822	.9718105
lprgnp	.2852386	.089022	3.20	0.003	.1041219	.4663553
t	-.0266633	.0048954	-5.45	0.000	-.036623	-.0167036
_cons	-6.663432	1.303295	-5.11	0.000	-9.315005	-4.011859

  - You need to `tsset` first to let Stata know it is time series data

Standard Errors Robust to Serial Correlation

- PRMINWGE.DTA: Effect of (log) minimum wage (*lmincov*) on employment rate (*lprepop*)

```
. newey lprepop lmincov lusgnp lprgnp t, lag(2)

Regression with Newey-West standard errors      Number of obs     =         38
maximum lag: 2                                  F(  4,        33) =      37.84
                                                Prob > F          =     0.0000

------------------------------------------------------------------------------
             |             Newey-West
     lprepop |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     lmincov |  -.2122612    .0457187    -4.64   0.000    -.3052766   -.1192459
      lusgnp |   .4860463    .2791124     1.74   0.091    -.0818122    1.053905
      lprgnp |   .2852386    .0996361     2.86   0.007     .0825275    .4879497
           t |  -.0266633    .0057558    -4.63   0.000    -.0383736    -.014953
       _cons |  -6.663432    1.536433    -4.34   0.000    -9.789329   -3.537535
------------------------------------------------------------------------------
```

Outline

1. Introduction: Relaxing the Independence Assumption

2. Clustered Data and Intra-Cluster Correlation

3. Time Series Data and Serial Correlation
   - Standard Errors Robust to Serial Correlation
   - Examples of Weakly Dependent Time Series*

Examples of Weakly Dependent Time Series

- As we saw, HAC s.e. require weak dependence of errors: $\gamma_j = \mathbb{E}[u_t u_{t+j}] \to 0$ as $j \to \infty$
- Here are two typical examples:
  - **First-order moving average process, MA(1)**:

$$u_t = e_t + \rho e_{t-1}, \quad e_t \text{ is i.i.d.}$$

$$\mathbb{E}[e_t] = 0, \quad \mathbb{V}[e_t] = \sigma^2, \quad \mathbb{E}[e_t e_{t-j}] = 0 \text{ for } j \neq 0$$

  - Then, mean, variance and covariance of u_t are

$$\mathbb{E}[u_t] = 0, \quad \mathbb{V}[u_t] = \sigma^2(1+\rho^2),$$

$$\mathbb{E}[u_t u_{t-1}] = \mathbb{E}[e_t e_{t-1}] + \rho\mathbb{E}[e_{t-1}^2] + \rho\mathbb{E}[e_t e_{t-2}] + \rho^2\mathbb{E}[e_{t-1}e_{t-2}] = \rho\sigma^2$$

$$\mathbb{E}[u_t u_{t-j}] = 0, \text{ for } j \geq 2$$

Examples of Weakly Dependent Time Series

- As we saw, HAC s.e. require weak dependence of errors: $\gamma_j = \mathbb{E}[u_t u_{t+j}] \to 0$ as $j \to \infty$
- Here are two typical examples:
  - **First-order autoregressive process, AR(1)**:

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

$$\mathbb{E}[e_t] = 0, \quad \mathbb{V}[e_t] = \sigma^2, \quad \mathbb{E}[e_t e_{t-j}] = 0 \text{ for } j \neq 0$$

  - Then, mean, variance and covariance of u_t are (not required)

$$\mathbb{E}[u_t] = 0, \quad \mathbb{V}[u_t] = \frac{\sigma^2}{1-\rho^2},$$

$$\mathbb{E}[u_t u_{t-j}] = \rho^j \times \frac{\sigma^2}{1-\rho^2}$$