1

Serial Correlation in Time Series Regressions

Hüseyin Taştan[1]

[1]Yıldız Technical University
Department of Economics

Econometrics II

2

Serial Correlation and Heteroskedasticity in Time Series Regressions

- In our previous classes, we learned that when the dynamics of a model is correctly specified, the errors will not be serially correlated.
- However, static and finite distributed lag models often have serially correlated errors even if there is no underlying misspecification of the model.
- Therefore, it is important to know the consequences and remedies for **serial correlation** for these useful classes of models.

3

Lecture Plan

- The properties of OLS when the errors contain serial correlation
- Testing for serial correlation
- Correcting for serial correlation under the assumption of strictly exogenous explanatory variables x
- Generalized Least Squares, Feasible GLS
- Heteroskedasticity in time series regression (ARCH, GARCH models) (Part II)

4

The properties of OLS with serially correlated errors

- The OLS estimator is unbiased under the first three Gauss-Markov assumptions for time series regressions (TS.1 through TS.3).
- As long as the explanatory variables are strictly exogenous, OLS estimators are unbiased, regardless of the degree of serial correlation in the errors.
- Heteroskedasticity in the errors does not cause bias in OLS estimators.
- In Chapter 11, we relaxed the strict exogeneity assumption and showed that, when the data are weakly dependent, OLS estimators are consistent (although not necessarily unbiased). This result does not depend on any assumption about serial correlation in the errors.

Efficiency and Inference

- Since the Gauss-Markov theorem (Theorem 10.4) requires both homoskedasticity and serially uncorrelated errors, OLS is no longer BLUE in the presence of serial correlation.
- So the usual OLS standard errors and test statistics are not valid, even asymptotically.
- We can see this by computing the variance of the OLS estimator under the first four Gauss-Markov assumptions and the AR(1) model for the error terms.

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

$$|\rho| < 1 \tag{2}$$

Efficiency and Inference

- e_t is an uncorrelated random variable with mean zero and constant variance. (also note the stability condition)
- Consider the variance of the slope estimator of β_1 in the simple regression:

$$y_t = \beta_0 + \beta_1 x_t + u_t$$

- For simplicity, assume that the sample average of the x_t is zero, $\bar{x} = 0$. Then the OLS estimator $\hat{\beta}_1$ of β_1 can be written as

$$\hat{\beta}_1 = \beta_1 + SST_x^{-1} \sum_{t=1}^{n} x_t u_t \tag{3}$$

where $SST_x = \sum_{t=1}^{n} x_t^2$. Now, to compute the variance of $\hat{\beta}_1$ (conditional on X), we must account for the serial correlation in the u_t.

Variance of the slope estimator

$$\begin{aligned}
\text{Var}(\hat{\beta}_1) &= SST_x^{-2} \text{Var}\left(\sum_{t=1}^{n} x_t u_t\right) \\
&= SST_x^{-2} \left(\sum_{t=1}^{n} x_t^2 \text{Var}(u_t) + 2 \sum_{t=1}^{n-1} \sum_{j=1}^{n-t} x_t x_{t+j} \text{E}(u_t u_{t+j})\right) \\
&= \frac{\sigma^2}{SST_x} + 2\left(\frac{\sigma^2}{SST_x^2}\right) \sum_{t=1}^{n-1} \sum_{j=1}^{n-t} \rho^j x_t x_{t+j} \qquad (4)
\end{aligned}$$

where $\sigma^2 = \text{Var}(u_t)$. Note that we use $\text{E}(u_t u_{t+j}) = \text{Cov}(u_t, u_{t+j}) = \rho^j \sigma^2$.
The first term in the equation (4) σ^2 / SST_x, is the variance of $\hat{\beta}_1$ when $\rho = 0$, which is the familiar OLS variance under the Gauss-Markov assumptions.

Efficiency and Inference

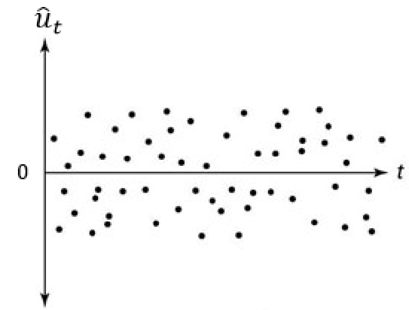
- The first term in equation (4) is the standard OLS variance under Gauss Markov assumptions when $\rho = 0$.
- If we ignore the serial correlation and estimate the variance in the usual way, the variance estimator will usually be biased when ($\rho \neq 0$) because it ignores the second term in equation (4).
- As we will see through later examples, $\rho > 0$ is most common for economic time series. Further, the independent variables in regression models are often positively correlated over time.
- So the usual OLS variance formula underestimates the true variance of the OLS estimator.

Efficiency and Inference

- The standard errors of $\hat{\beta}$s in the presence of serial correlation is invalid.
- Therefore, t statistics are no longer valid for testing single hypotheses.
- Since a smaller standard error means a larger t statistic (remember the formula for t test), the usual t statistics will often be too large when $\rho > 0$.
- Additionally, the usual F and LM statistics for testing multiple hypotheses are also invalid.
- To sum up, OLS estimator is inefficient and the usual statistical inference procedures are invalid.

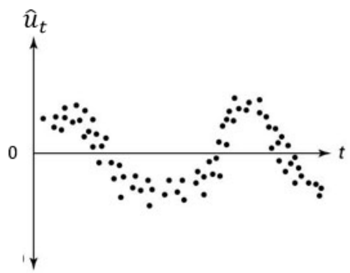
How to detect serial correlation?

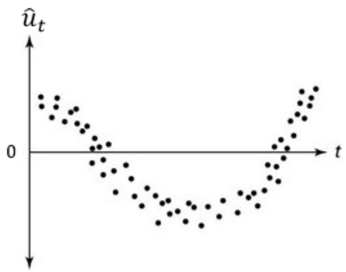
- One way to detect the existence of serial correlation is to estimate the model by OLS and then plot the residuals against time.
- If there there is no serial correlation we expect the residuals to be distributed randomly across time, without any detectible pattern.
- The following graph illustrates residuals without serial correlation.



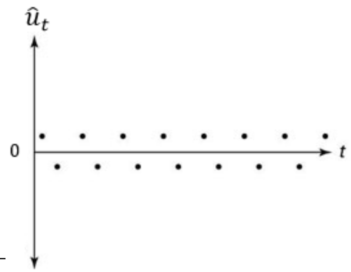
How to detect serial correlation?

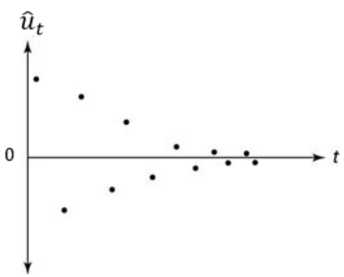
Positive Autocorrelation:





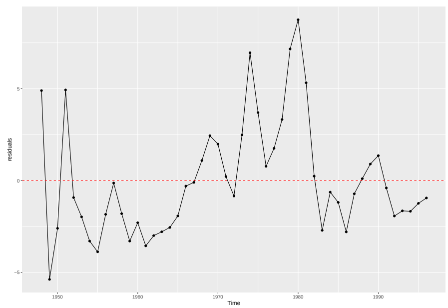
Negative Autocorrelation:





Example: Static Phillips Curve

Regression of the inflation rate on the unemployment rate produces the following residuals:



Inspection of this graph suggests that positive values are generally followed by positive values, and vice versa. This indicates that there may be serial correlation in the residuals. But we need to conduct formal hypothesis tests to diagnose serial correlation.

Testing for Serial Correlation (Autocorrelation)

- We first consider the case when the regressors are **strictly exogenous**. Recall that this requires the error to be uncorrelated with the regressors in all time periods.
- **So, it rules out models with lagged dependent variables.**

u_t follows AR(1) and xs are strictly exogenous:

- Although there are many ways in which the error terms can be serially correlated, the most popular model (and the simplest to work with) is the AR(1) model in equations (1) and (2).
- We will learn two tests: (1) t test and (2) Durbin-Watson (DW) test.
- In both tests, we assume that the error term follows an AR(1) process:

$$u_t = \rho u_{t-1} + e_t, \quad e_t \; White\ Noise(0, \sigma_e^2)$$

where $|\rho| < 1$

t test with Strictly Exogenous Regressors

- Under the assumption that the explanatory variables are strictly exogenous: the expected value of u_t, given the entire history of independent variables, is zero. In addition, we must assume that:

$$E(e_t|u_{t-1}, u_{t-2}, ...) = 0 \tag{5}$$

$$Var(e_t|u_{t-1}) = Var(e_t) = \sigma_e^2 \tag{6}$$

- The null hypothesis states that there is no autocorrelation:

$$H_0 : \rho = 0. \tag{7}$$

- The alternative can be two-sided or one-sided. For economic time series, $H_1 : \rho > 0$, i.e., positive autocorrelation, generally makes more sense.

t test for AR(1) Serial Correlation with Strictly Exogenous Regressors

1. Run the OLS regression of y_t on $x_{t1}, ..., x_{tk}$ and obtain the OLS residuals, $\hat{u}_t$, for all $t = 1, 2, ..., n$.
2. Run the regression of

$$\hat{u}_t \quad \text{on} \quad \hat{u}_{t-1} \quad , \quad t = 2, ..., n \tag{8}$$

3. Obtain the coefficient $\hat{\rho}$ on $\hat{u}_{t-1}$ and its t statistic, $t_{\hat{\rho}}$.
4. Use $t_{\hat{\rho}}$ to test $H_0 : \rho = 0$ against $H_0 : \rho \neq 0$ in the usual way. (Actually, since $\rho > 0$ is often expected a priori, the alternative can be $H_0 : \rho > 0$).
5. Typically, we conclude that serial correlation is a problem to be dealt with only if H_0 is rejected at the 5 percent level. We can also use and report the p-value for the test.

Example: Static Phillips Curve

- For the static Phillips curve (see ch.10), the test regression (in step 2) yields $\hat{\rho} = 0.573$ and $t = 4.93$, $p - value = 0.000$ (with 48 observations). See the R output below.
- This is very strong evidence of positive, first order serial correlation. This implies that the standard errors and t statistics are not valid.

```
> ttestreg <- dynlm(residual.s ~ L(residual.s))
> coeftest(ttestreg)
   t test of coefficients:

               Estimate Std. Error t value  Pr(>|t|)
(Intercept)    -0.11340    0.35940 -0.3155    0.7538
L(residual.s)   0.57297    0.11613  4.9337 1.098e-05
```

Note: `residual.s` is the set of residuals from the static regression. See the R Lab05 notes.

Example: Expectations Augmented Phillips Curve

- We also estimated a particular expectations augmented Phillips curve using the same data (see ch.11).
- By contrast, the test for AR(1) serial correlation in the expectations augmented curve gives $\hat{\rho} = -0.036$ and $t = -0.287$, $p - value = 0.775$ (with 47 observations): there is no evidence of AR(1) serial correlation in the expectations augmented Phillips curve.

```
> reg.ea <- dynlm( d(inf) ~ unem, data=tsdata, end=1996)
> residual.ea <- resid(reg.ea)
> coeftest( dynlm(residual.ea ~ L(residual.ea)) )
   t test of coefficients:

                Estimate Std. Error t value Pr(>|t|)
(Intercept)     0.194166   0.300384  0.6464   0.5213
L(residual.ea) -0.035593   0.123891 -0.2873   0.7752
```


The Durbin-Watson Test Under Classical Assumptions

- Another test for AR(1) serial correlation is the **Durbin-Watson test**. **The Durbin- Watson (DW) statistic** is also based on the OLS residuals:

$$DW = \frac{\sum_{t=2}^{n} (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{n} \hat{u}_t^2} \quad (9)$$

- It can easily be shown that DW and $\hat{\rho}$ from (13) are closely linked:

$$DW \approx 2(1 - \hat{\rho}) \quad (10)$$

- One reason this relationship is not exact is that $\hat{\rho}$ has $\sum_{t=2}^{n} \hat{u}_{t-1}^2$ in its denominator, while the DW statistic has the sum of squares of all OLS residuals in its denominator. But even with small samples the approximation is often very good. Therefore, tests based on DW and the t test based on $\hat{\rho}$ are conceptually the same.

Durbin-Watson Test

- Durbin and Watson (1950), derive the distribution of DW (conditional on X), something that requires the full set of classical linear model assumptions, including normality of the error terms.
- We assume that an intercept is included in the original model.
- There must be no lagged values of the dependent variable as regressors, $y_{t-1}, y_{t-2}, \ldots$ (remember the strict exogeneity assumption).
- Under these assumptions, the sampling distribution of the DW statistic is nonstandard: there are two critical values in the test's decision rule.
- **Upper (dU)** and **lower(DL)** bounds for the critical values that depend on the desired significance level, the alternative hypothesis, **the number of observations** n, and **the number of explanatory variables** (k).

Durbin-Watson Test

- Usually, the DW test is computed for the alternative

$$H_1 : \rho > 0 \quad (11)$$

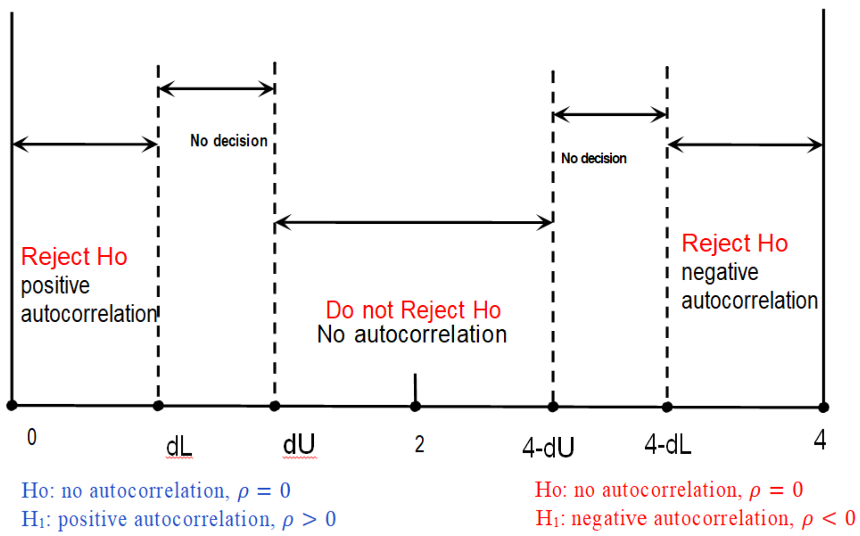
- From the approximation in (10), $\hat{\rho} \approx 0$ implies that $DW \approx 2$, and $\hat{\rho} > 0$ implies that $DW < 2$.
- When $0 < DW < 2$, this generally implies the existence of positive autocorrelation but we need to test this formally. Also $2 < DW < 4$ may indicate negative serial correlation.
- There are two sets of critical values. These are usually labeled as dU (for upper) and dL (for lower).

Durbin-Watson Test Decision Rules

- If $DW < dL$, then we reject H_0 in favor of the alternative (positive autocorrelation);
- if $DW > dU$, we fail to reject H_0 (no autocorrelation).
- If $dL \leq DW \leq dU$, the test is **inconclusive**.
- Critical values are found using simulation techniques and tabulated for reference.
- For example, for $n = 50$ and $k = 2$ 5% critical values are $d_L = 1.4625$, $d_U = 1.6283$. So, for this case, any DW value less than 1.4625 would indicate positive autocorrelation.
- Alternatively, p-value of the DW statistic can be found using simulation techniques.

See the illustrative graph for the DW decision rules in the next slide. Note that the inconclusive region may not be that wide as indicated in the graph.

Durbin-Watson Test



Durbin-Watson Test 5% Critical Values

	k = 1		k = 2		k = 3		k = 4		k = 5	
n	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
30	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83
31	1.36	1.50	1.30	1.57	1.23	1.65	1.16	1.74	1.09	1.83
32	1.37	1.50	1.31	1.57	1.24	1.65	1.18	1.73	1.11	1.82
33	1.38	1.51	1.32	1.58	1.26	1.65	1.19	1.73	1.13	1.81
34	1.39	1.51	1.33	1.58	1.27	1.65	1.21	1.73	1.15	1.81
35	1.40	1.52	1.34	1.58	1.28	1.65	1.22	1.73	1.16	1.80
36	1.41	1.52	1.35	1.59	1.29	1.65	1.24	1.73	1.18	1.80
37	1.42	1.53	1.36	1.59	1.31	1.66	1.25	1.72	1.19	1.80
38	1.43	1.54	1.37	1.59	1.32	1.66	1.26	1.72	1.21	1.79
39	1.43	1.54	1.38	1.60	1.33	1.66	1.27	1.72	1.22	1.79
40	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79
45	1.48	1.57	1.43	1.62	1.38	1.67	1.34	1.72	1.29	1.78
50	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77
55	1.53	1.60	1.49	1.64	1.45	1.68	1.41	1.72	1.38	1.77
60	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73	1.41	1.77
65	1.57	1.63	1.54	1.66	1.50	1.70	1.47	1.73	1.44	1.77
70	1.58	1.64	1.55	1.67	1.52	1.70	1.49	1.74	1.46	1.77
75	1.60	1.65	1.57	1.68	1.54	1.71	1.51	1.74	1.49	1.77
80	1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74	1.51	1.77
85	1.62	1.67	1.60	1.70	1.57	1.72	1.55	1.75	1.52	1.77
90	1.63	1.68	1.61	1.70	1.59	1.73	1.57	1.75	1.54	1.78
95	1.64	1.69	1.62	1.71	1.60	1.73	1.58	1.75	1.56	1.78
100	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76	1.57	1.78

n=sample size, k=number of explanatory variables (excluding constant)

DW Test Example: Phillips Curves

Let's compute the DW test statistic for the static Phillips Curve model:

```
> # DW test for the static Phillips curve
> library(lmtest)
> dwtest(reg.s)
 Durbin-Watson test
  data:  reg.s
  DW = 0.8027, p-value = 7.552e-07
```

From the table, we see that the critical values are $d_L = 1.5$ and $d_U = 1.59$ at 5% significance level.
Because $DW = 0.8027 < 1.5$ we reject the null hypothesis of no serial correlation ($\rho = 0$) in favor of the positive autocorrelation $\rho > 0$. Also notice the p-value is very small implying that the null can be rejected more decisively.

DW Test Example: Phillips Curves

Now, compute the DW test statistic for the expectations augmented Phillips Curve:

```
> # DW test for the expectations-augmented Phillips curve
  data:  reg.ea
  DW = 1.7696, p-value = 0.1783
```

For the expectations augmented Phillips curve $DW = 1.77$, approximately. And this larger than d_U.
Thus, we fail to reject the null hypothesis. Residuals seem to be serially uncorrelated in this case. Also the p value is 0.18 so we don't reject the null cannot even at 10% level.

Testing for AR(1) Serial Correlation without Strictly Exogenous Regressors

- **When the explanatory variables are not strictly exogenous**, so that one or more x_{tj} is correlated with u_{t-1}, neither the t test nor the Durbin-Watson test statistic are valid, even in large samples .
- The leading case of **nonstrictly exogenous regressors** occurs when the model contains a lagged dependent variable: y_{t-1} and u_{t-1}. are obviously correlated.
- Durbin's alternative test is used when there are any number of non-strictly exogenous explanatory variables:

Testing for Serial Correlation with General Regressors

1. Run the OLS regression of y_t on $x_{t1}, ..., x_{tk}$ and obtain the OLS residuals, $\hat{u}$, for all $t = 1, 2, ..., n$.
2. Run the regression of

$$\hat{u}_t \quad \text{on} \quad x_{t1}, ..., x_{tk}, \hat{u}_{t-1} \quad , \quad t = 2, ..., n \tag{12}$$

3. Obtain the coefficient $\hat{\rho}$ on $\hat{u}_{t-1}$ and its t statistic, $t_{\hat{\rho}}$.
4. Use $t_{\hat{\rho}}$ to test $H_0 : \rho = 0$ against $H_0 : \rho \neq 0$ in the usual way (or use a one-sided alternative).

- In the second step, we regress the OLS residuals on all independent variables, including an intercept, and the lagged residual. The t statistic on the lagged residual is a valid test.

Example 12.2: Testing for AR(1) Serial Correlation in the Minimum Wage Equation

- In this example (from ch.10), we assume that the underlying stochastic processes are weakly dependent, but we allow them to contain a linear time trend. Letting $\hat{u}_t$ denote the OLS residuals, we run the regression of

$$\hat{u}_t \quad \text{on} \quad log(mincov_t), log(prgnp_t), log(usgnp_t), t, \hat{u}_{t-1} \quad \text{on},$$

  using 37 observations. Note that we include all x variables in the test regression.
- The estimated coefficient on $\hat{u}_{t-1}$ is $\hat{\rho} = 0.481$ with $t = 2.89$, p-value$= 0.007$. Therefore, there is strong evidence of AR(1) serial correlation in the errors, which means the t statistics for the $\hat{\beta}_j$ that we obtained before are not valid for inference.

Testing for Higher Order Serial Correlation

- The previous test can easily extended to higher orders of serial correlation. For example, in the AR(2) model

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + e_t$$

- suppose that we wish to test the null hypothesis:

$$H_0 : \rho_1 = 0, \rho_2 = 0 \tag{13}$$

- This alternative model of serial correlation allows us to test for second order serial correlation. As always, we estimate the model by OLS and obtain the OLS residuals, $\hat{u}_t$. Then, we can run the regression of

$$\hat{u}_t \quad \text{on} \quad x_{t1}, ..., x_{tk}, \hat{u}_{t-1}, \hat{u}_{t-2} \quad , \quad t = 3, ..., n$$

Testing for Higher Order Serial Correlation

- To obtain the F test for joint significance of $\hat{u}_{t-1}$ and $\hat{u}_{t-2}$. If these two lags are jointly significant at a small enough level, say 5 percent, then we reject the null hypothesis and conclude that the errors are serially correlated.
- More generally, we can test for serial correlation in the autoregressive model of order q:

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + ... + \rho_q u_{t-q} + e_t \tag{14}$$

- The null hypothesis is

$$H_0 : \rho_1 = 0, \rho_2 = 0, ..., \rho_q = 0 \tag{15}$$

Steps of Testing for AR(q) Serial Correlation

1. Run the OLS regression of y_t on $x_{t1}, ..., x_{tk}$ and obtain the OLS residuals, $\hat{u}$, for all $t = 1, 2, ..., n$.
2. Run the regression of

$$\hat{u}_t \quad \text{on} \quad x_{t1}, ..., x_{tk}, \hat{u}_{t-1}, \hat{u}_{t-2}, ..., \hat{u}_{t-q} \quad , \quad t = (q+1), ..., n \tag{16}$$

3. Compute the F test for joint significance of $\hat{u}_{t-1}$, $\hat{u}_{t-2}$,...,$\hat{u}_{t-q}$.

- The test requires the homoskedasticity assumption:

$$Var(u_t | x_t, u_{t-1}, ..., u_{t-q}) = \sigma^2 \tag{17}$$

LM Test for AR(q) Serial Correlation

- An alternative to computing the F test is to use the Lagrange multiplier (LM) form of the statistic. (We covered the LM statistic for testing exclusion restrictions in Chapter 5 for cross-sectional analysis.) The LM statistic for testing no serial correlation is simply:

$$LM = (n - q) R^2_{\hat{u}} \tag{18}$$

- where $R^2_{\hat{u}}$ is just the usual R-squared from the test regression (in the second step). Under the null hypothesis, LM statistic follows χ^2_q. This is usually called the **Breusch-Godfrey test for AR(q) serial correlation**. The LM statistic also requires homoscedasticity, but it can be made robust to heteroskedasticity.

Correcting for Serial Correlation with Strictly Exogenous Regressors

- When the errors u_t follow the AR(1) model and our variables are strictly exogenous, we can correct the test statistics.
- For simplicity, consider the case with a single explanatory variable:

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

- Since the problem in this equation is serial correlation in the u_t, it makes sense to transform the equation to eliminate the serial correlation. For $t \geq 2$, we write

$$y_{t-1} = \beta_0 + \beta_1 x_{t-1} + u_{t-1}$$

$$y_t = \beta_0 + \beta_1 x_t + u_t$$

Correcting for Serial Correlation

- Now, if we multiply this first equation by ρ and subtract it from the second equation, we get

$$y_t - \rho y_{t-1} = (1-\rho)\beta_0 + \beta_1(x_t - \rho x_{t-1}) + e_t, \ t \geq 2$$

where we have used the fact that $e_t = u_t - \rho u_{t-1}$.
- We can write this as

$$\tilde{y}_t = (1-\rho)\beta_0 + \beta_1 \tilde{x}_t + e_t, \ t \geq 2 \qquad (19)$$

where

$$\tilde{y}_t = y_t - \rho y_{t-1}, \tilde{x}_t = x_t - \rho x_{t-1} \qquad (20)$$

are called the **quasi-differenced data**.
- When $\rho = 1$, these are differenced data (not quasi anymore), but remember we are assuming $\rho < 1$. The error terms in (19) are serially uncorrelated; in fact, this equation satisfies all of the Gauss-Markov assumptions.

Correcting for Serial Correlation

- If we knew ρ, we could estimate β_0 and β_1 by regressing $\tilde{y}_t$ on $\tilde{x}_t$, provided we divide the estimated intercept by $(1-\rho)$.
- Adding more regressors changes very little. For $t \geq 2$, we use the equation

$$\tilde{y}_t = (1-\rho)\beta_0 + \beta_1 \tilde{x}_{t1} + ... + \beta_k \tilde{x}_{tk} + e_t, \qquad (21)$$

where $\tilde{x}_{tj} = x_{tj} - \rho x_{t-1,j}$.
- For $t = 1$, we have $\tilde{y}_1 = (1-\rho^2)^{0.5} y_1$, $\tilde{x}_{1j} = (1-\rho^2)^{0.5} x_{1j}$ and the intercept is $(1-\rho^2)^{0.5}\beta_0$.

Generalized Least Squares (GLS)

- For a given ρ, it is easy to transform the data and to run OLS.
- **GLS estimation** = Transforming the model as outlined and then using OLS to estimate the transformed model
- Unless $\rho = 0$, the GLS estimator will generally be different from the original OLS estimator.
- The GLS estimator is BLUE, and, since the errors in the transformed equation are serially uncorrelated and homoskedastic, t and F statistics from the transformed equation are at least asymptotically valid if e_t is normally distributed.
- To be able to apply GLS in case of AR(1) errors, we need to know the true value of ρ which is generally unknown in practice. Estimating ρ first and applying GLS $\Rightarrow$ Feasible GLS (FGLS).

Feasible GLS Estimation with AR(1) Errors

- The problem with the GLS estimator is that ρ is rarely known in practice.
- Consistent estimation of ρ: regress the OLS residuals on their lagged counterparts.
- Next, we use this estimate, $\hat{\rho}$, in place of ρ to obtain the quasi-differenced variables.
- We then use OLS on the equation

$$\tilde{y}_t = \beta_0 \tilde{x}_{t0} + \beta_1 \tilde{x}_{t1} + ... + \beta_k \tilde{x}_{tk} + error_t, \qquad (22)$$

  where $\tilde{x}_{t0} = (1 - \hat{\rho})$ for $t \geq 2$, and $\tilde{x}_{10} = (1 - \hat{\rho})^{0.5}$.
- This results in the **feasible GLS** (FGLS) estimator of the β_j.

Steps in Feasible GLS Estimation with AR(1) Errors

1. Run the OLS regression of y_t on $x_{t1}, ..., x_{tk}$ and obtain the OLS residuals, $\hat{u}_t$, $t = 1, 2, ..., n$.
2. Run the regression $\hat{u}_t$ on $\hat{u}_{t-1}$ and obtain $\hat{\rho}$.
3. Apply OLS to equation (22) to estimate $\beta_0, \beta_1, ..., \beta_k$. The usual standard errors, t statistics, and F statistics are asymptotically valid.

FGLS estimator is consistent and statistical inference procedures are valid asymptotically.

Feasible GLS Estimation with AR(1) Errors

- FGLS estimator is not unbiased, so it is not BLUE.
- But it is asymptotically more efficient than the OLS estimator when the AR(1) model for serial correlation holds, the explanatory variables are strictly exogenous, and all the time series are weakly dependent.
- In the econometrics litarature, there are several names for FGLS estimation of the AR(1) model that depend on different methods of estimating ρ and different treatment of the first observation.
- **Cochrane-Orcutt (CO) estimation** omits the first observation in estimating $\hat{\rho}$.
- **Prais-Winsten (PW) estimation** uses the first observation in the previously suggested way. Asymptotically, these two methods are equivalent but they may differ in small samples.

Example 12.5: Static Phillips Curve

Comparison of OLS and Prais-Winsten (PW) estimation:

TABLE 12.2 Dependent Variable: *inf*

Coefficient	OLS	Prais-Winsten
unem	.468 (.289)	−.716 (.313)
intercept	1.424 (1.719)	8.296 (2.231)
$\hat{\rho}$	———	.781
Observations	49	49
R-squared	.053	.136

See the R lab for estimation details. The Cochrane-Orcutt (CO) estimation yields -0.67 on the unem variable. This example highlights that the OLS and FGLS (PW or CO) can be markedly different. Here, because PW result is consistent with the inflation-unemployment tradeoff, we may ignore the OLS estimates and focus on the FGLS results. Notice that the AR(1) coefficient on the residuals is fairly high, $\hat{\rho} = 0.781$

Differencing and Serial Correlation

- Consider the simple regression model with AR(1) errors:

$$y_t = \beta_0 + \beta_1 x_t + u_t, \quad u_t \sim AR(1), \quad t = 1, 2, ..., n \tag{23}$$

- The usual OLS inference procedures are invalid if y_t and x_t are integrated of order one, or I(1), e.g. random walk.
- If the error term u_t follows a random walk process, i.e., $u_t = u_{t-1} + e_t$, the equation above makes no sense because, among other things, the variance of u_t grows with t (nonstationary). In this case, we need to difference the equation:

$$\Delta y_t = \beta_1 \Delta x_t + \Delta u_t, \quad t = 2, ..., n \tag{24}$$

- If u_t follows a random walk, then $e_t \equiv \Delta u_t$ has zero mean, a constant variance, and is serially uncorrelated. Thus, assuming that e_t and Δx_t are uncorrelated, we can estimate the differenced model above by OLS, where we lose the first observation.

Serial Correlation-Robust Standard Errors

- We derived the heteroskedasticity-robust standard errors for cross-sectional regression.
- In a similar fashion, we can compute the serial correlation-robust standard errors of OLS estimators.
- In fact, it is possible to make standard errors robust to both serial correlation and heteroskedasticity.
- In the literature, these are known as Heteroskedasticity and Autocorrelation Consistent (or, HAC) standard errors (see the discussion in Section 12.5 in the text)