Serial Correlation and Heteroskedasticity in Time Series Regressions,
Wooldridge (2013), Chapter 12

- Serially Correlated Errors: Consequences
- Testing for Serial Correlation
- Generalised Least squares (GLS) with strictly exogenous regressors
- Serial Correlation-Robust Standard Errors
- Heteroskedasticity in Time Series Regressions
  - Autoregressive Conditional Heteroskedasticity
- GLS with heteroskedasticity and serial correlation.

- With assumptions TS.1 through TS.3, OLS estimators are *unbiased*

Recall:

Assumption (TS.1 - linearity in parameters)

The stochastic process $\{(y_t, x_{t1}, x_{t2}, ..., x_{tk}); t = 1, 2, ..., n\}$ *follows the linear model:*

$$y_t = \beta_0 + \beta_1 x_{t1} + ... + \beta_k x_{tk} + u_t$$

Assumption (TS.2 - no perfect collinearity)

No regressor independent variable is a constant nor a perfect linear combination of the other regressors.

Let X denotes the matrix of all regressors for all time-periods, i.e. $X = [x_{tj}; t = 1, ..., n \;\&\; j = 1; ...; k]$.

Assumption (TS.3 - zero conditional mean)

$$E(u_t|X) = 0, t = 1, 2, \ldots, n$$

- With assumptions TS.1′ through TS.3′, OLS estimators are *consistent*

Assumption (TS.1′ - linearity in parameters)

The stochastic process $\{(y_t, x_{t1}, x_{t2}, ..., x_{tk}); t = 1, 2, ..., n\}$ *is stationary and weakly dependent and follows the linear model:*

$$y_t = \beta_0 + \beta_1 x_{t1} + ... + \beta_k x_{tk} + u_t$$

Assumption (TS.2′ - no perfect collinearity)

No regressor independent variable is a constant nor a perfect linear combination of the other regressors.

Write $\mathbf{x}_t = (x_{t1,\ldots,}x_{tk})$.

Assumption (TS.3′ - zero conditional mean)

$$E(u_t|\mathbf{x}_t) = 0, t = 1, 2, \ldots, n$$

- In Time Series we often have serial correlation in the errors (TS.5 or TS.5′ is violated): For each $t; s = 1, 2, ..., n$ such that $t \neq s$:

$$Corr(u_t, u_s | \mathbf{x}_t, \mathbf{x}_s) \neq 0$$

- But for inference, results are*NOT valid* if TS.5 (or TS.5′) fail: With serial correlation in the errors, usual OLS variances are NOT valid.

- If we do not have lagged dependent variables as regressors, and have serial correlation, OLS is unbiased and consistent but the usual formulas for the standard errors are *not valid*.
- Serial correlation might lead to inconsistency if we have lagged dependent variables as regressors, *but not always*.

Example 1:

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

where $E[u_t | y_{t-1}] = 0$ and $\{u_t\}$ are serially correlated. In this case **OLS is consistent**.

Example 2:

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

and

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

$t = 2, \ldots, n$, where e_t are i.i.d., $|\rho| < 1$ and $E[e_t | u_{t-1}, u_{t-2}, \ldots] = E[e_t | y_{t-1}, y_{t-2}, \ldots] = 0$.

Example 2: (cont)

- Then,

$$\begin{aligned} Cov(y_{t-1}, u_t) &= E[y_{t-1}(\rho u_{t-1} + e_t)] \\ &= \rho E(y_{t-1} u_{t-1}) \\ &= \rho E[y_{t-1}(y_{t-1} - \beta_0 - \beta_1 y_{t-2})] \neq 0 \end{aligned}$$

  unless $\rho = 0$.
- In this case the *OLS estimators are not consistent* for β_0 , β_1 . This is a special form of autocorrelation.

Models with lagged dependent variables and serial correlation in the errors can often be easily transformed into models without serial correlation in the errors

Example 2 (cont): Notice that:

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

where $E[e_t|y_{t-1}, y_{t-2}, \ldots] = 0$ and
$E[y_t|y_{t-1}, y_{t-2}, \ldots] = E[y_t|y_{t-1}, y_{t-2}] = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + e_t$.

- Thus, the *"relevant"* model is an $AR(2)$ model for y. With further conditions on the parameters (that ensure stability) we can estimate the α_j's consistently.
- Hence, if you have serial correlation you can add a lagged dependent variable to the model and that might lead to a model with no serial correlation.

- Want to be able to test for whether the errors are serially correlated or not.
- Consider the multiple regression model

$$y = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + u.$$

- Want to test the null that $\rho = 0$ in

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

  $t = 2, \ldots, n$, where u_t is the model error term and e_t is iid.
- With *strictly exogenous regressors* $E(u_t|X) = 0, t = 1, 2, \ldots, n$, the test is: straightforward:

1. Obtain the OLS residuals of the original model.
2. simply regress the residuals on (one period) lagged residuals (use OLS).
3. Use a typical t-test for $H_0 : \rho = 0$ (don't need intercept). (The asymptotic distribution of the t-statistic is standard normal).

- We assume TS.1 through TS.4 (homoskedasticity) hold. If TS.4 fails use a heteroskedasticity robust statistic as in chapter 8.

This test detects correlation between u_t and u_{t-1} but not between u_t and u_{t-2}.

- **Remark:** This test is only valid if we do not have lagged dependent variables as regressors.

Example: US data: 1948-1996

$$Inflation_t = \beta_0 + \beta_1 Unemployment_t + u_t$$

Dependent Variable: INF

Method: Least Squares

Sample: 1 49

Included observations: 49

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.42361	1.719015	0.828154	0.4118
UNEM	0.467626	0.289126	1.617376	0.1125

R-squared	0.052723	Mean dependent var	4.108163
Adjusted R-squared	0.032568	S.D. dependent var	3.182821
S.E. of regression	3.130562	Akaike info criterion	5.160262
Sum squared resid	460.6198	Schwarz criterion	5.237479
Log likelihood	-124.426	F-statistic	2.615904
Durbin-Watson stat	**0.8027**	Prob(F-statistic)	0.11249

Example: Regress residuals of previous equation on past residuals

Dependent Variable: RESID01
Method: Least Squares
Sample(adjusted): 2 49
Included observations: 48 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RESID01(-1)	0.572735	0.115013	4.979738	0

R-squared	0.344633	Mean dependent var	-0.10207
Adjusted R-squared	0.344633	S.D. dependent var	3.046154
S.E. of regression	2.466005	Akaike info criterion	4.66369
Sum squared resid	285.8156	Schwarz criterion	4.702673
Log likelihood	-110.929	Durbin-Watson stat	1.351045

What do you conclude at 5% level?

- An alternative is the *Durbin-Watson (DW) statistic*, which is calculated by many packages. Need TS.1 to TS.6 holding:

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

- Notice that if we do not have serial correlation $\hat{\rho} \approx 0$. Thus $DW \approx 2$.
- If the DW statistic is around 2, then we do not reject H_0 (absence of serial correlation of AR(1) type), while if it is significantly < 2 we reject the null hypothesis (against the alternative $H_1 : \rho > 0$, the most typical).
- Critical values are difficult to calculate as they will depend on the regressors of the model, making the t test easier to work with.

- If the regressors are *not strictly exogenous*, both tests for serial correlation described before will not work.
- For example: If there are *lagged dependent variables* as regressors then neither previous t- statistic or DW test for serial correlation will work.
- There are alternative tests that are valid under TS.1′ through TS.2′ (that are valid also with TS.1 through TS.4).
- Regress the residual (or y) on the *lagged residual* and all of the x's.
- Then test the significance of the *lagged residual* using the usual t-statistic.

Example: (cont) Regress residuals of previous equation on past residuals and regressors

Dependent variable: RESID

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.705871	1.464288	1.847909	0.0711
UNEM	-0.47356	0.24753	-1.913156	0.062
RESID(-1)	0.659484	0.125004	**5.27569**	0
R-squared	0.376972	Mean dependent var		-1.97E-16
Durbin-Watson stat	1.818217	Prob(F-statistic)		0.000019

Can test for AR(q) serial correlation in the same basic manner as AR(1):

$$\begin{aligned} y &= \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + u, \\ u_t &= \rho_1 u_{t-1} + \rho_2 u_{t-2} + ... + \rho_q u_{t-q} + e_t \end{aligned}$$

- where e_t is an iid sequence with $E(e_t|u_{t-1}, ..., u_{t-q}) = 0$, $Var(e_t|u_{t-1}, ..., u_{t-q}) = \sigma_e^2$
- The null is $H_0 : \rho_1 = ... = \rho_q = 0$. The alternative is H_1 : the is one ρ_j such that $\rho_j \neq 0$.

Steps:

1. Regress the residual (or y) on the q **lagged residual** and all of the x's.
2. Then use the usual test of multiple restrictions (on the coefficients of the lagged residuals).
3. Can use F test or LM test, where the LM version is called a *Breusch-Godfrey test* and is $LM = (n-q)R^2 \overset{a}{\sim} \chi^2(q)$ using R^2 from residual regression
4. If there is heteroskedasticity can use heteroskedastic robust statistics.

Example: Alternative Specification – Augmented Phillips Curve

$$Inflation_t - Inflation_{t-1} = \beta_0 + \beta_1 (Unemployment_t - NaturalRate^*) + u_t$$

Dependent Variable: INF-INF(-1)				
Method: Least Squares				
Sample(adjusted): 2 49				
Included observations: 48 after adjusting endpoints				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.030581	1.37681	2.201161	0.0328
UNEM	-0.54259	0.230156	-2.357475	0.0227
R-squared	0.107796	Mean dependent var		-0.10625
Adjusted R-squared	0.0884	S.D. dependent var		2.566926
S.E. of regression	2.450843	Akaike info criterion		4.671515
Sum squared resid	276.3051	Schwarz criterion		4.749482
Log likelihood	-110.116	F-statistic		5.557689
Durbin-Watson stat	**1.769648**	Prob(F-statistic)		0.02271

Regress residuals of previous equation on lagged residuals and regressors

Dependent variable: RESID

C	-0.83592	1.31912	-0.633697	0.5296
UNEM	0.144109	0.220873	0.652453	0.5175
RESID(-1)	-0.05971	0.138511	-0.431046	0.6685
RESID(-2)	-0.4168	0.140903	-2.958062	0.005
R-squared	0.16695	Mean dependent var		4.39E-16

- If there is serial correlation in the errors, inference using the "usual" formulas with OLS is not valid.
- What can we do?
- As in the cross-sectional case with Heteroskedasticity we have 2 alternatives:
  1. Use *Generalised Least Squares* (in case of *strictly exogenous regressors*).
  2. Use OLS but compute *Serial-Correlation Robust Standard errors* (in correctly specified models satisfying the *contemporaneous exogeneity* assumption).

- Let us assume TS.1 through TS.4 (homoskedasticity) hold in our model, but TS.5 (no serial correlation) fails to hold. Also, assume stationarity and weak dependence as in TS.1′.

Consider the model

$$y_t = \beta_0 + \beta_1 x_{t1} + ... + \beta_k x_{tk} + u_t$$

Assume errors follow an AR(1) so

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

$t = 2, \ldots, n$ where $|\rho| < 1$ and e_t is an i.i.d. sequence with $E[e_t|X] = 0$ and also $Var[e_t|X] = \sigma_e^2$

$$Var(u_t|X) = \frac{\sigma_e^2}{1 - \rho^2}$$

Let us first assume that ρ is known.
We can transform the equation so we have no serial correlation in the errors

$$y_t = \beta_0 + \beta_1 x_{t1} + ... + \beta_k x_{tk} + u_t$$

Notice that

$$\rho y_{t-1} = \rho\beta_0 + \rho\beta_1 x_{t-1,1} + ... + \rho\beta_k x_{t-1,k} + \rho u_{t-1}$$

Hence

$$\begin{aligned} y_t - \rho y_{t-1} &= (1-\rho)\beta_0 + \beta_1(x_{t1} - \rho x_{t-1,1}) + ... + \beta_k(x_{tk} - \rho x_{t-1,k}) + e_t, \\ \text{for } t &\geq 2 \end{aligned}$$

- In this quasi-differenced model for $t \geq 2$, TS.1 through TS.5 hold!

So far we ignored the first observation. Can transform equation for $t = 1$,

$$y_1 = \beta_0 + \beta_1 x_{11} + ... + \beta_k x_{1k} + u_1$$

so that TS.1 through TS.5 hold for $t = 1$ (not yet the case since u_1 and e_t have a different variance):

$$Var(u_1|X) = \sigma_e^2/(1-\rho^2)$$

How can we do it?
We can multiply equation by $(1-\rho^2)^{1/2}$ to have TS.1 through TS.5 holding for $t = 1$

$$\begin{aligned}(1-\rho^2)^{1/2}y_1 &= (1-\rho^2)^{1/2}\beta_0 + (1-\rho^2)^{1/2}\beta_1 x_1 + ... + (1-\rho^2)^{1/2}\beta_k x_{1k} \\ &\quad +(1-\rho^2)^{1/2}u_1\end{aligned}$$

In this quasi-differenced model, TS.1 through TS.5 hold!

The transformed model is:

$$\tilde{y}_t = \beta_0 \tilde{x}_{t0} + \beta_1 \tilde{x}_{t1} + ... + \beta_k \tilde{x}_{tk} + \tilde{e},$$

where

$$\tilde{y}_t = \begin{cases} (1-\rho^2)^{1/2} y_1 & t = 1 \\ y_t - \rho y_{t-1} & t \geq 2 \end{cases}, \tilde{x}_{t0} = \begin{cases} (1-\rho^2)^{1/2} & t = 1 \\ 1-\rho & t \geq 2 \end{cases},$$

and

$$\tilde{x}_{tj} = \begin{cases} (1-\rho^2)^{1/2} x_{1,j} & t = 1 \\ x_{t,j} - \rho x_{t-1,j} & t \geq 2 \end{cases}., j = 1, ..., k$$

- If is ρ known, can estimate the transformed regression by OLS.
- This is called the GLS estimator of the original model.
- GLS is *BLUE* if TS.1 through TS.5 hold in the transformed model.
- Can use t and F tests from the transformed equation to conduct inference on the parameters of the original equation.
- These tests are valid (asymptotically) if TS.1 through TS.5 hold in the transformed model (along with stationarity and weak dependence in the original variables) and distributions (conditional on X) are exact (and with minimum variance) if TS.6 holds for the e_t.

- Problem: don't know ρ, need to get an estimate first
- Run OLS on the original model and then regress residuals $\hat{u}_t$ on lagged residuals $\hat{u}_{t-1}$ (with OLS). The obtained estimator $\hat{\rho}$ is the estimator of ρ.
- Finally, replace ρ by $\hat{\rho}$ and use OLS on the transformed regression. This is a *feasible GLS* (FGLS) estimator (or *Prais-Winsten estimator*).
- If we ignore the first equation ($t = 1$) we have the *Cochrane-Orcutt estimator* (also a FGLS estimator).
- These FGLS estimators are not unbiased, but are consistent if TS.1 through TS.5 hold in the transformed model (along with stationarity and weak dependence in the original model).
- There is no difference, asymptotically (for large sample sizes), between the two procedures.

- t and F tests from the transformed equations are valid (asymptotically). If TS.6 holds for the e_t , relevant distributions also hold only asymptotically.
- FGLS is asymptotically more efficient than OLS
- This basic method can be extended to allow for higher order serial correlation,AR(q), in the error term. Econometrics packages deal automatically with estimation of AR serial correlation

Example:

$$Inflation_t = \beta_0 + \beta_1 Unemployment_t + u_t$$

US data: 1948-1996

Dependent Variable: INF
Method: Least Squares
Sample: 1 49
Included observations: 49

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.42361	1.719015	0.828154	0.4118
UNEM	0.467626	0.289126	1.617376	0.1125

R-squared	0.052723	Mean dependent var	4.108163
Adjusted R-squared	0.032568	S.D. dependent var	3.182821
S.E. of regression	3.130562	Akaike info criterion	5.160262
Sum squared resid	460.6198	Schwarz criterion	5.237479
Log likelihood	-124.426	F-statistic	2.615904
Durbin-Watson stat	**0.8027**	Prob(F-statistic)	0.11249

Need to estimate ρ so we regress residuals of previous equation on past residuals

Dependent Variable: RESID01
Method: Least Squares
Sample(adjusted): 2 49
Included observations: 48 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RESID01(-1)	0.572735	0.115013	4.979738	0

R-squared	0.344633	Mean dependent var	-0.10207
Adjusted R-squared	0.344633	S.D. dependent var	3.046154
S.E. of regression	2.466005	Akaike info criterion	4.66369
Sum squared resid	285.8156	Schwarz criterion	4.702673
Log likelihood	-110.929	Durbin-Watson stat	1.351045

Transform variables with estimated and apply OLS to transformed equation, t≥2

Dependent Variable: INF-.572735*INF(-1)				
Method: Least Squares				
Sample(adjusted): 2 49				
Included observations: 48 after adjusting endpoints				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
1-.572735	5.509824	2.037318	2.70445	0.0096
UNEM-.572735*UNEM(-1)	-0.27931	0.321911	-0.867665	0.3901
R-squared	0.016103	Mean dependent var		1.658889
Adjusted R-squared	-0.00529	S.D. dependent var		2.349715
S.E. of regression	2.355918	Akaike info criterion		4.592512
Sum squared resid	255.316	Schwarz criterion		4.670478
Log likelihood	-108.22	Durbin-Watson stat		1.217816

Comparison of OLS and Cochrane-Orcutt

Coefficient	**OLS**	**Cochrane-Orcutt**
		ρ^=.573
Intercept	**1.42**	**5.51**
(s.e.)	(1.72)	(2.04)
{t}	{.83}	{2.71}
Unem	**0.47**	**-0.28**
(s.e.)	(0.289)	(0.32)
{t}	{1.62}	{-0.869}
(Observations)	49	48
R-Squared	**0.0527**	**0.016**

- FGLS asymptotic distributions rely on strict exogeneity of the regressors
- If strict exogeneity does not hold, we can still use OLS (it is consistent with only contemporaneous exogeneity, or TS.1′, along with TS.2′ and TS.3′). Further, we can calculate *serial correlation-robust standard errors*
- Actually, with only TS.1′ through TS.3′ we can derive *Heteroskedasticity and Serial Correlation robust standard errors*, also known as *Heteroskedasticity and Autocorrelation (HAC) robust standard errors* or *Newey-West HAC standard errors*.

- *Heteroskedasticity and serial correlation-robust standard errors*:

$$se(\hat{\beta}_j) = \left[\text{“}se(\hat{\beta}_j)\text{”}/\hat{\sigma}\right]^2 \sqrt{\hat{v}}$$

The usual OLS standard errors are normalized and then „inflated" by a correction factor.

where

- "$se\left(\hat{\beta}_j\right)$"$= \frac{\hat{\sigma}}{\sqrt{SST_j(1-R_j^2)}}$ $(SST_j = \sum_{i=1}^{n}\left(x_{ij} - \bar{x}_j\right)^2$ and R_j^2 is the R^2 from the regressing x_j on all other $x's$.)
- $\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} \hat{u}_i^2}{n-k-1} = \frac{SSR}{df}$.
- **Remark:** these formulae are the formulae of the standard error or $\hat{\beta}_j$ and the estimator of the variance of u_t proposed under the Gauss Markov assumptions.

- Correction factor for serial correlation (*Newey-West formula*):

$$\hat{v} = \sum_{t=1}^{n} \hat{a}_t^2 + 2 \sum_{h=1}^{g} [1 - h/(g+1)] \left(\sum_{t=h+1}^{n} \hat{a}_t \hat{a}_{t-h} \right)$$

$$\hat{a}_t = \hat{r}_t \, \hat{u}_t$$

This term is the product of the residuals and the residuals of a regression of x_{tj} on all other explanatory variables

- The integer g is called *lag-truncation* or *bandwidth* and controls how much serial correlation is taken into account:
  - $g = 1$:
    $$\hat{v} = \sum_{t=1}^{n} \hat{a}_t^2 + \sum_{t=2}^{n} \hat{a}_t \hat{a}_{t-1}$$
  - $g = 2$:

The weight of higher order autocorrelations is declining

$$\hat{v} = \sum_{t=1}^{n} \hat{a}_t^2 + (4/3) \sum_{t=2}^{n} \hat{a}_t \hat{a}_{t-1} + (2/3) \sum_{t=3}^{n} \hat{a}_t \hat{a}_{t-2}$$

- t-statistics, Wald statistics and LM statistics are computed using these standard errors (Econometrics packages compute this).

Discussion of serial correlation-robust standard errors

- For the integer g, values such as g=1 or g=2 are normally sufficient (there are more involved rules of thumb for how to choose g)
- Serial correlation-robust standard errors are only valid asymptotically; they may be severely biased if the sample size is not large enough
- The bias is the higher the more autocorrelation there is; if the series are highly correlated, it might be a good idea to difference them first
- Serial correlation-robust errors should be used if there is *serial correlation* and *contemporaneous exogeneity holds*.

Example:

Dependent Variable : INF   Observations: 49

Newey-West HAC Standard Errors & Covariance (lag truncation=3)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.42361	1.515019	0.939665	0.3522
UNEM	0.467626	0.291606	1.603624	0.1155
R-squared	0.052723	Mean dependent var		4.108163
Adjusted R-squared	0.032568	S.D. dependent var		3.182821
S.E. of regression	3.130562	Akaike info criterion		5.160262
Sum squared resid	460.6198	Schwarz criterion		5.237479
Log likelihood	-124.4264	F-statistic		2.615904
Durbin-Watson stat	0.8027	Prob(F-statistic)		0.11249

Effects on OLS: Similar consequences to those of serial correlation and to those of conditional heteroskedasticity in cross-section regressions:

- OLS estimates of coefficients remain consistent.
- Usual OLS standard errors are invalid
- Usual t-test and F-tests are invalid

Dealing with Conditional Heteroskedasticity
As in cross-section regressions we can go one of two routes:

1. Compute corrected standard errors and implement corrected t-test and F-tests. (HAC standard errors are valid here).
2. Compute Generalised least squares estimates.

The usual tests of conditional heteroskedasticity can be applied in time-series regressions but are only valid in the absence of serial correlation. In contrast we can implement versions of the tests for serial correlation that are valid in the presence of conditional heteroskedasticity.

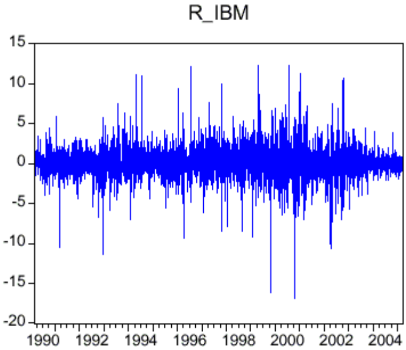
- Consider the following static model:

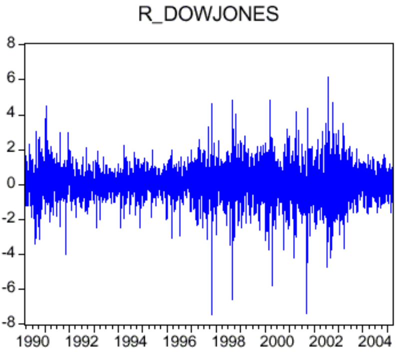
$$y_t = \beta_0 + \beta_1 x_t + u_t,$$

  such that $E(u_t|X) = 0$,where X denotes all n outcomes of x_t

- Assume that the Gauss-Markov assumptions hold. This means that the OLS estimators are BLUE. The homoskedasticity assumption says that $Var(u_t|X)$ is constant, .
- Even if the variance of u_t given X is constant, there are other ways that heteroskedasticity can arise.
- *Autoregressive Conditional Heteroskedasticity* is a special type of Heteroskedasticity.
- Robert Engle was awarded the 2003 Nobel Memorial Prize in Economic Sciences specially for introducing this topic.

Empirical Evidence: Daily financial returns display *volatility clustering*: periods of high volatility alternate with more tranquil periods.
Example: Daily log-returns on IBM stock price and Dow Jones index, March 1990 – March 2005





- The 1rst model of this type was the Autoregressive Conditional Heteroskedasticity (ARCH) model.
- The 1rst order ARCH model assumes that

$$E(u_t^2|u_{t-1}, u_{t-2}, ...) = \alpha_0 + \alpha_1 u_{t-1}^2, \quad (1)$$

where we leave the conditioning on X implicit.

- If $E(u_t|u_{t-1}, u_{t-2}, ...) = 0$ (no serial correlation in u_t) this implies that

$$\text{var}(u_t|u_{t-1}, u_{t-2}, ...) = \alpha_0 + \alpha_1 u_{t-1}^2$$

- Since conditional variances must be positive, this model only makes sense if $\alpha_0 > 0$.and $\alpha_1 \geq 0$; if $\alpha_1 = 0$, there are no dynamics in the variance equation.

- It also implies that

$$u_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + v_t$$

where $E(v_t | u_{t-1}, u_{t-2}, ...) = 0$. This equation looks like an autoregressive model in u_t^2 (hence the name ARCH). The stationary condition for this equation is $\alpha_1 < 1$, just as in the usual AR(1) model (as $\alpha_1 \geq 0$).

- What implications does (1) have for OLS?
  - Because we began by assuming the Gauss-Markov assumptions hold, OLS is *BLUE*.
  - Even if u_t is not normally distributed, we know that the usual OLS test statistics are asymptotically valid under Assumptions TS.1′ through TS.5′, which are satisfied by static and distributed lag models with *ARCH* errors.

- If OLS still has desirable properties under ARCH, why should we care about ARCH forms of heteroskedasticity in static and distributed lag models?
    - It is possible to get *consistent* (but not unbiased) estimators of the β_j that are asymptotically more *efficient* than the OLS estimators. WLS, based on estimating (1), will do the trick.
    - Economists and financial analysts have become interested in dynamics in the conditional variance. Since variance is often used to measure volatility, and volatility is a key element in asset pricing theories, ARCH models have become important in empirical finance.

Let us now consider an autoregressive distributed lag model:

$$E(y_t|x_t, y_{t-1}, x_{t-1}, y_{t-2}, \ldots) = \beta_0 + \beta_1 x_t + \beta_2 y_{t-1} + \beta_3 x_{t-1}$$

- Define $u_t = y_t - E(y_t|x_t, y_{t-1}, x_{t-1}, y_{t-2}, ...)$, therefore

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 y_{t-1} + \beta_3 x_{t-1} + u_t$$

- The 1rst order ARCH model in this case assumes that

$$\begin{aligned}\text{var}(y_t|x_t, y_{t-1}, x_{t-1}, y_{t-2}, ...) &= \text{var}(u_t|x_t, y_{t-1}, x_{t-1}, y_{t-2}, ...) \\ &= \alpha_0 + \alpha_1 u_{t-1}^2,\end{aligned}$$

In this model:

- OLS is *consistent*.
- The homoscedasticity assumption TS.4′ is necessarily violated as $\text{var}(u_t|x_t, y_{t-1}, x_{t-1}, y_{t-2}, ...) =$
$= \alpha_0 + \alpha_1 \left(y_{t-1} - \beta_0 - \beta_1 x_{t-1} - \beta_2 y_{t-2} - \beta_3 x_{t-2}\right)^2,$
- In this case, *heteroscedasticity-robust standard error and test statistics* should be computed, or a *FGLS/WLS*-procedure should be applied
- Using a *FGLS/WLS*-procedure will also increase efficiency

- Consider again the autoregressive distributed lag model.
- Usually one finds that more lags are needed to explain the conditional variance, leading to the ARCH(q) model:

$$\text{var}(u_t | x_t, y_{t-1}, x_{t-1}, y_{t-2}, ...) = \alpha_0 + \alpha_1 u_{t-1}^2 + \ldots + \alpha_q u_{t-q}^2,$$

which is equivalent to

$$E(u_t^2 | x_t, y_{t-1}, x_{t-1}, y_{t-2}, ...) = \alpha_0 + \alpha_1 u_{t-1}^2 + \ldots + \alpha_q u_{t-q}^2.$$

- Testing for ARCH effects: Run the regression of y on the regressors and compute the residuals: $\hat{u}_t$
- Lagrange-Multiplier (LM) test against ARCH, which is based on $LM = (n - q) \times R^2 \sim \chi^2(q)$. R^2 is the R^2 of the regression

$$\hat{u}_t^2 = \gamma_0 + \gamma_1 \hat{u}_{t-1}^2 + \ldots + \gamma_q \hat{u}_{t-q}^2 + v_t.$$

- The null hypothesis is H_0 : (no ARCH effects) $\alpha_1 = ... = \alpha_q = 0$ vs H_1 : There is one $\gamma_j \neq 0$.

Example: Testing for ARCH-effects in stock returns

$$return_t = \beta_0 + \beta_1 return_{t-1} + u_t$$

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

$$\Rightarrow \; u_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + v_t$$

$$\underset{}{\hat{u}_t^2} = \underset{(0.44)}{2.95} + \underset{(.036)}{.337}\ \hat{u}_{t-1}^2$$

$$n = 688, R^2 = .1136$$

Can have violation of TS.4 (Homoskedasticity) and TS.5 (No serial correlation) simultaneously. Assume still that TS.1 through TS.3 hold (along with stationarity and weak dependence).
Consider the model

$$
\begin{aligned}
y_t &= \beta_0 + \beta_1 x_{t1}, + ... + \beta_k x_{tk} + u_t, \\
u_t &= v_t \sqrt{h_t}, \\
v_t &= \rho v_{t-1} + e_t.
\end{aligned}
$$

where X are independent of e_t for all t,h_t is a function of the regressors and $|\rho| < 1$, and the process $\{e_t\}$ has zero mean and constant variance σ_e^2 and is serially uncorrelated.

Notice that

$$Var(u_t|X) = h_t\sigma_v^2.$$

where $\sigma_v^2 = var(v_t|X) = \sigma_e^2/(1-\rho^2)$.
Therefore

$$\frac{y_t}{\sqrt{h_t}} = \frac{\beta_0}{\sqrt{h_t}} + \beta_1\frac{x_{t1}}{\sqrt{h_t}}, +\ldots + \beta_k\frac{x_{tk}}{\sqrt{h_t}} + v_t$$

- We can estimate the function h exactly as in chapter 8 of Wooldridge (2013) [chapter 7 of the program]:
  - Run the regression of y_t on an intercept, $x_{t1}, ...x_{tk}$ and save the residuals $\hat{u}_t$.
  - Regress $\log(\hat{u}_t^2)$ on an intercept, $x_{t1}, \ldots, x_{tk}$ and obtain the fitted values $\widehat{\log(\hat{u}_t^2)}$.
  - Obtain the estimates of h_t: $\hat{h}_t = \exp\left(\widehat{\log(\hat{u}_t^2)}\right)$.
- Estimate the transformed equation

$$\frac{y_t}{\sqrt{\hat{h}_t}} = \frac{\beta_0}{\sqrt{\hat{h}_t}} + \beta_1 \frac{x_{t1}}{\sqrt{\hat{h}_t}}, + ... + \beta_k \frac{x_{tk}}{\sqrt{\hat{h}_t}} + error_t$$

  by Cochrane-Orcutt or Prais-Winsten estimators.
- This leads to a feasible GLS estimator that is asymptotically efficient. Test statistics from Cochrane-Orcutt or Prais-Winsten are asymptotically valid.