AUTOCORRELATION

[or Serial Correlation]

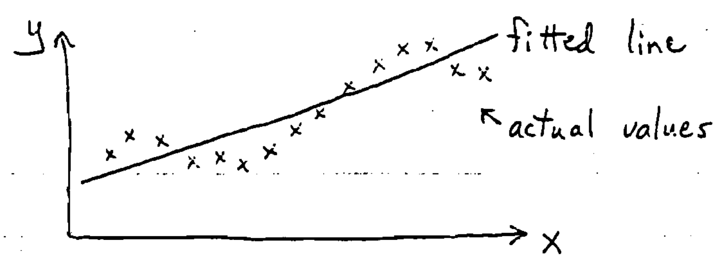
The problem: $Cov(\epsilon) = \sigma^2 \Omega$

where Ω is not diagonal.

i.e. $Cov(\epsilon_t, \epsilon_s) \neq 0$ for $t \neq s$

Autocorrelation is usually associated with time series data; the disturbance term in this period is correlated with disturbances of previous periods.
[What happens randomly today is correlated with what happened randomly yesterday.]

Graphically:



If the residual was positive (neg.) last period, it is likely to be positive (neg.) again this period.

→ Disturbances not independent;
Ω not diagonal.

OLS Results:

i) $\hat{\beta}$ unbiased, consistent, but inefficient
ii) s^2 biased & inconsistent.
iii) t's, F's are wrong.

The cure is GLS!
So find Ω^{-1}; find a V; get $\tilde{\beta}$.

The usual assumption about Ω:

First Order Autoregressive Process –

$$\epsilon_t = \rho \epsilon_{t-1} + u_t$$

where $\underline{|\rho| < 1}$ (stability requiremt)

and the u_t are iid $N(0, \sigma_u^2)$.

First Order Autocorrelation assumes that the error term is composed of two parts – one that satisfies the ideal conditions, and one that is some fraction of the previous error term.

Note: If $\rho = 0$, $\epsilon_t = u_t$ & $\nexists$ no problem.

Notation: Let $Var(u_t) = \sigma_u^2$;
$Var(\epsilon_t) = \sigma_\epsilon^2$.

Handout

Lemma ①:

$$\epsilon_t = u_t + \rho u_{t-1} + \rho^2 u_{t-2} + \cdots$$
$$= \sum_{j=0}^{\infty} \rho^j u_{t-j}$$

Also,

$\epsilon_t = u_t + \rho u_{t-1} + \cdots + \rho^{i-1} u_{t-i+1} + \rho^i u_{t-i} + \rho^{i+1} u_{t-i-1} + \cdots$

$\epsilon_{t-i} = u_{t-i} + \rho u_{t-i-1} + \cdots$

$Cov(\epsilon_t, \epsilon_{t-i})$

$= Cov(\rho^i u_{t-i}, u_{t-i}) + Cov(\rho^{i+1} u_{t-i-1}, \rho u_{t-i-1}) + \cdots$

$= \rho^i \sigma_u^2 + \rho^{i+2} \sigma_u^2 + \cdots$

$= \sigma_u^2 (\rho^i, \rho^{i+2}, \cdots)$

$= \sigma_u^2 \dfrac{\rho^i}{1-\rho^2}$

$= \rho^i \sigma_\epsilon^2$

☆

Proof:

$$\begin{aligned}
\epsilon_t &= \rho \epsilon_{t-1} + u_t \\
&= \rho[\rho \epsilon_{t-2} + u_{t-1}] + u_t \\
&= \rho^2 \epsilon_{t-2} + u_t + \rho u_{t-1} \\
&= \rho^2[\rho \epsilon_{t-3} + u_{t-2}] + u_t + \rho u_{t-1} \\
&= \rho^3 \epsilon_{t-3} + u_t + \rho u_{t-1} + \rho^2 u_{t-2}
\end{aligned}$$

After n substitutions;

$$= \underbrace{\rho^n \epsilon_{t-n}}_{\left[\substack{\text{as } n \to \infty \\ \text{this} \to 0}\right]} + \underline{\underline{\sum_{j=0}^{n-1} \rho^j u_{t-j}}}$$

QED

<u>Note:</u> u_t is uncorrelated with <u>previous</u> ϵ's (ϵ_{t-i}, i
$\rightarrow Cov(u_t, \epsilon_{t-1}) = Cov(u_t, \text{past } u_t\text{'s}) = 0$.

Lemma ②:

$$\sigma_\epsilon^2 = \frac{\sigma_u^2}{1-\rho^2}$$

Proof:

$$\epsilon_t = u_t + \rho u_{t-1} + \rho^2 u_{t-2} + \cdots$$
$$Var(\epsilon_t) = \sigma_u^2 + \rho^2 \sigma_u^2 + \rho^4 \sigma_u^2 + \cdots$$

(since the terms are independent)

$$= \sigma_u^2 (1 + \rho^2 + (\rho^2)^2 + (\rho^2)^3 + \cdots)$$

(geometric series with common ratio, ρ^2)

$$= \sigma_u^2 \left(\frac{1}{1-\rho^2}\right)$$

QED.

Again, $\epsilon \sim N(0, \sigma^2\Omega)$;

$Cov(\epsilon) = \sigma^2 \Omega$ -- we're finding Ω under the ass. of First Order Autocorrelation. Then we can find V & get $\tilde{\beta}$.

Thm : $Cov(\epsilon_t, \epsilon_{t-i}) = \sigma_\epsilon^2 \rho^i \qquad i \geq 0$.

Proof: $Cov(\epsilon_t, \epsilon_{t-i}) = E[\epsilon_t * \epsilon_{t-1}]$

$$= E[(\rho^i \epsilon_{t-i} + u_t + \rho u_{t-1} + \rho^2 u_{t-2} + \cdots + \rho^{i-1} u_{t-i+1}) * \epsilon_{t-i}]$$

$\Uparrow$

Substituting i times as in Lemma 1.

$$= E[\rho^i \epsilon_{t-i} * \epsilon_{t-i}]$$

Since any u is uncorrelated with previous ϵ

$$= \rho^i \, Var(\epsilon_{t-i})$$
$$= \sigma_\epsilon^2 \rho^i$$

QED

Since the ϵ_t are iid $N(0, \sigma_\epsilon^2 \Omega)$ where the <u>diagonal</u> of Ω is all 1's

Note: This Thm $\longrightarrow$ the correlation between ϵ_t and the ϵ in previous periods <u>decreases exponentially</u> as it is compared with ϵ's further back in time.

$\longrightarrow$ ρ^i decreases exponentially as i $\uparrow$es.

$$E(\epsilon_t^2) = \sigma_\epsilon^2$$
$$E(\epsilon_t \epsilon_{t-1}) = \rho \sigma_\epsilon^2$$
$$E(\epsilon_t \epsilon_{t-2}) = \rho^2 \sigma_\epsilon^2$$
$$E(\epsilon_t \epsilon_{t-3}) = \rho^3 \sigma_\epsilon^2$$
$$\vdots$$

Thus $$Cov(\epsilon) = \sigma_\epsilon^2 \Omega = \sigma_\epsilon^2 \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{T-1} \\ \rho & 1 & \rho & \cdots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{T-3} \\ & \vdots & & & \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \cdots & 1 \end{bmatrix}$$

<u>Note:</u> $\sigma_\epsilon^2 \Omega_{ij} = \sigma_\epsilon^2 \rho^{|i-j|}$

or $Cov(\epsilon_t, \epsilon_s) = \sigma_\epsilon^2 \rho^{|t-s|}$.

$\longrightarrow$ The Covariance depends upon
how far apart the obs. are
in the time series.
The further apart,
the smaller the Covariance.

Sensible ! Random events further back in time have <u>less</u> influence on the current disturbance than more recent random events.

<u>Note:</u> <u>Stationary</u> scheme ; the covariance does not depend on where you are in the scheme. [$Cov(\epsilon, \epsilon) = Cov(\epsilon, \epsilon$

Claim ①: s^2 (from OLS) is usually an <u>underestimate</u> of σ^2 if;
(i) $\exists$ a constant term, and
(ii) $\rho > 0$.

(See Schmidt)

Claim ②: $(X'X)^{-1}$ is usually <u>smaller</u> than $(X'X)^{-1} X'\Omega X (X'X)^{-1}$ if;
i) $\rho > 0$, and
ii) the regressors are "positively correlated" over time.
--- typical, with time trends.

Note: OLS t-ratios $\rightarrow \dfrac{\hat{\beta}_i}{\sqrt{s^2 (X'X)^{-1}_{ii}}}$

Claim ① $\rightarrow$ s^2 too small.
Claim ② $\rightarrow$ $(X'X)^{-1}_{ii}$ too small.

$\Rightarrow$ t-ratios usually <u>overstated</u>!

Hence you have false confidence in estimates.
— open yourself up for criticism.

67 xerox

The cure, GLS.

Recall $\text{Cov}(\epsilon) = \sigma_\epsilon^2 \Omega = \sigma_\epsilon^2 \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{T-1} \\ \rho & 1 & \rho & \cdots & \rho^{T-2} \\ & & \vdots & & \end{bmatrix}$

Lemma ③:

$$\Omega^{-1} = \frac{1}{1-\rho^2} \begin{bmatrix} 1 & -\rho & 0 & 0 & \\ -\rho & (1+\rho^2) & -\rho & 0 & \cdots \\ 0 & -\rho & (1+\rho^2) & -\rho & \\ 0 & 0 & -\rho & \ddots & \\ & \vdots & & & \end{bmatrix}$$

Proof: $\Omega^{-1}\Omega = I$.
Verify yourself.

Lemma ④: $\Omega^{-1} = V'V$

where $V = \frac{1}{\sqrt{1-\rho^2}} \begin{bmatrix} \sqrt{1-\rho^2} & 0 & 0 & 0 & \\ -\rho & 1 & 0 & 0 & \\ 0 & -\rho & 1 & 0 & \cdots \\ 0 & 0 & -\rho & 1 & \\ & \vdots & & & \end{bmatrix}$

Proof: $V'V = \Omega^{-1}$.
Verify yourself.

Now we can construct GLS $\tilde{\beta}$.

68
xerox

2 ways to construct $\tilde{\beta}$:

(i) $\tilde{\beta} = (X'\Omega^{-1}X)^{-1} X'\Omega^{-1} Y$ (use Lemma 3)

(ii) Regress VY on VX (use Lemma 4)

The Transformation; $(VY) = (VX)\beta + (V\epsilon)$

$$VY = \begin{bmatrix} \sqrt{1-\rho^2}\ y_1 \\ y_2 - \rho y_1 \\ y_3 - \rho y_2 \\ \vdots \\ y_T - \rho y_{T-1} \end{bmatrix}$$

$$VX = \begin{bmatrix} \sqrt{1-\rho^2}\ x_{11} & \sqrt{1-\rho^2}\ x_{12} & & \sqrt{1-\rho^2}\ x_{1K} \\ x_{21} - \rho x_{11} & x_{22} - \rho x_{12} & \cdots & x_{2K} - \rho x_{1K} \\ x_{31} - \rho x_{21} & x_{32} - \rho x_{22} & & x_{3K} - \rho x_{2K} \\ \vdots & \vdots & & \vdots \\ x_{T1} - \rho x_{T-1,1} & x_{T2} - \rho x_{T-1,2} & & x_{TK} - \rho x_{T-1,K} \end{bmatrix}$$

$$V\epsilon = \begin{bmatrix} \sqrt{1-\rho^2}\ \epsilon_1 \\ \epsilon_2 - \rho \epsilon_1 \\ \epsilon_3 - \rho \epsilon_2 \\ \vdots \\ \epsilon_T - \rho \epsilon_{T-1} \end{bmatrix}$$

Called Cochrane - Orcutt Transformation

$$(y_t - \rho y_{t-1}),\ \{(x_{t1} - \rho x_{t-1,1})\ (x_{t2} - \rho x_{t-1,2})\ \ldots\}$$

$$t = 2, 3, \ldots, T$$

The first (top) observation is typically dropped. ⊛

Why C-O transformation works; (1) $\Rightarrow\ y_t = \alpha + \beta x_t + \epsilon_t$

(2) $\times\rho \Rightarrow\ y_{t-1} = \alpha + \beta x_{t-1} + \epsilon_{t-1}$

Since $\epsilon_t = \rho \epsilon_{t-1} + u_t$,

[subtract (2) from (1), voila

$$\epsilon_t - \rho \epsilon_{t-1} = u_t .$$

The transformed disturbance term is

$$V\epsilon = \begin{bmatrix} \sqrt{1-\rho^2}\ \epsilon_1 \\ u_2 \\ u_3 \\ \vdots \\ u_T \end{bmatrix} \quad \text{and} \quad V\epsilon \sim N(0, \sigma_u^2 I_T)$$

* Consider top term; $Var(\sqrt{1-\rho^2}\ \epsilon) = (1-\rho^2)\ \sigma_\epsilon^2 = \sigma_u^2$

[See Lemma 2.]

⊛ Dropping first obs. may be bad

(see Schmidt.)

Problem: Don't know ρ.
Need a consistent estimate, $\hat{\rho}$.

Then we'll have $\hat{\Omega} = f(\hat{\rho})$
is a consistent estimate of Ω,

and GLS $\tilde{\beta}^* = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}Y$
will be asymptotically efficient.
— [see previous notes]

[✩ Actually, just plug $\hat{\rho}$ into the
Cochrane - Orcutt transformation to get $\tilde{\beta}^*$.

Estimation of ρ

1) Totally lazy way; let $\rho = 1$.
[first differencing the data.]

2) Cochrane - Orcutt method

$$\text{let } \hat{\rho} = \frac{\sum_{t=2}^{T} e_t e_{t-1} / T-1}{\sum_{t=1}^{T} e_t^2 / T} \qquad (\tfrac{T}{T-1} \to 1)$$

where the e_t are the OLS residuals.

— almost the correlation coeff. of e_t & e_{t-1}.

— Regress e_t on e_{t-1} $\Rightarrow$ $\hat{\rho}$.

Could use iterative procedure here;

(i) Run OLS $\rightarrow \hat{\beta} \rightarrow e_t$

[these residuals are based on ineff. $\hat{\beta}_{OLS}$ estima

(ii) from $e_t \rightarrow \hat{\rho} \rightarrow \tilde{\beta}^*_{GLS}$.

⋮

Could keep going.
The residuals based on $\tilde{\beta}^*_{GLS}$
are "better";
get new $\hat{\rho}$, new $\tilde{\beta}^*$, . . .

Estimates of $\hat{\rho}$ & $\tilde{\beta}^*$
presumably get better
each iteration.

HOWEVER, the asymptotic properties
remain the same
– all asymptotically efficient.

3) Durbin method [JRSS, 1960]

Transformed Model; $(y_t - \rho y_{t-1}) = \beta_1 (X_{t1} - \rho X_{t-1,1}) + \cdots + u_t$

Then; $y_t = \rho y_{t-1} + \beta_1 x_{t1} - (\beta_1 \rho) x_{t-1,1} + \cdots + u_t$

Run this regression — (2K+1) regressors $\{\beta_i, -\rho\beta_i, \rho$

$\hat{\rho}$ = first coefficient.

4) Hildreth - Lu (almost MLE ; finding top of hill)

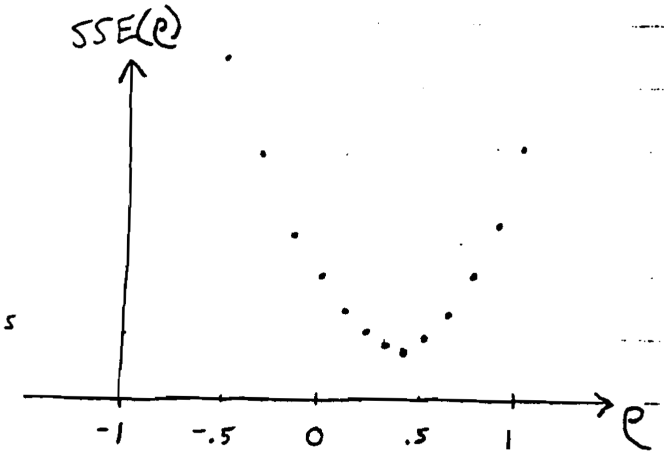
Transformed Model ;

$$(y_t - \rho\, y_{t-1}) = \beta_1 (x_{t1} - \rho\, x_{t-1,1}) + \cdots + \beta_K (x_{tK} - \rho\, x_{t-1,K}) + u_t$$

Run this regression for several ρ's between -1 & 1 and pick the one that minimizes $SSE(\rho)$.

– A search procedure.

Can then divide the 10th (.4 - .5) into hundredths, & run 10 more regressions to get a finer estimate of $\hat{\rho}$.



By minimizing $SSE(\rho) = \sum_{t=2}^{T} \left[(y_t - \rho\, y_{t-1}) - \sum_{j=1}^{K} \beta_j (x_{tj} - \rho\, x_{t-1,j}) \right]^2$

wrt β_i & ρ,
we are maximizing something very similar to the Likelihood Function.

$\rightarrow$ Consistent & asymptotically efficient.

<u>Note:</u> Hildreth - Lu <u>simultaneously</u> estimates ρ & the β_i.
This is intuitively appealling.

<u>BUT</u>, requires running many regressions.

If you're satisfied with a consistent & asymptotically efficient est., C - O is adequate.
<u>(and less work!)</u>

<u>Note:</u> If the iterated C - O method is continued until $\hat{\rho}$ converges,

the C-O $\hat{\rho}$ will = the H-L $\hat{\rho}$.

[The iterated C-O is a local min. for $SSE(\rho)$.]

[By all of these last 3 procedures, [②, ③, & ④] $\tilde{\beta}^*$ is <u>consistent</u> and <u>asymptotically efficient</u>.

i.e. With <u>large samples</u>,
These estimates are the best possible.

(better than $\hat{\beta}_{OLS}$)]

What about Small-Sample Properties?

Which method of estimating ρ is most efficient?

Monte Carlo (simulation) results

Kmenta & Gilbert, JASA, 1968, conduct Monte Carlo experiment.

Pick the following;

Model: $y_t = \alpha + \beta x_t + \epsilon_t$

where $\epsilon_t = \rho \epsilon_{t-1} + u_t$

Parameters: $\alpha = 10$

$\beta = 2$

$\sigma_u^2 = 1$

$\rho = .8$

Sample Size: $T = 20$

What we don't have is observations, So get them from a random # generator.

from Random # generator ;

$$\left.\begin{matrix} \omega_1 \\ \vdots \\ \omega_{20} \end{matrix}\right\} \text{ from iid } N(0,1)$$

Construct 1^{st} order autoregressive disturbance structure:

$$\epsilon_1 = \omega_1$$
$$\epsilon_2 = .8\,\epsilon_1 + \omega_2$$
$$\epsilon_3 = .8\,\epsilon_2 + \omega_3$$
$$\vdots$$
$$\epsilon_{20} = .8\,\epsilon_{19} + \omega_{20}$$

[Recall, $Var(\omega_t) = 1$, $\rho = .8$]

Pick the X's ; $X_1, \ldots, X_{20}$

Generate the y's from the model.

Then run the regression
estimate α, β, ρ, σ_ϵ^2
according to C-O, D, & H-L.

Repeat this procedure 100 times ;
Compare the C-O, D, & H-L estimates.

– examine which → true values better.

We can presumably then run the regression 100 <u>more</u> times with <u>different</u> values for the parameters, and again compare the results. etc.

This is an experimental method for deriving the small-sample properties (distribution & efficiency) of the various estimates.

<u>Kmenta & Gilbert results:</u>

From 50 runs

		OLS	H-L	C-O
$\tilde{\beta}^*$	Mean	2.0008	1.9995	1.9998
	Std Dev.	.0634	.0402	.0412

Recall true $\beta = 2$.

Note: All estimates appear to be unbiased.
H-L & C-O are clearly efficient relative to OLS.

$\Downarrow$

In this case, the asymptotic properties manifest themselves after about 20 observations!

i.e. with as few as 20 obs., GLS beats OLS.

Problems with Monte Carlo experiments

① Any results are dependent on the parameters picked ($\alpha, \beta, \rho, \sigma_\epsilon^2, T$).

Also you can generally only pick a few parameter values to examine, since the experimental procedure is exponential!

e.g. If you want to test
3α's, 3β's, $3\sigma_\epsilon^2$'s, 6ρ's, & $20\ T$'s,

$3^3 * 6 * 20$ = 3200 regressions to run!

② It is possible that the estimators may not have moments.
(the means & std. errors do not exist).

Depends on the random # generator.

$\Rightarrow$ the mean, variance of the random generator obs. is what?

Testing for Autocorrelation

Under assumption of
1st Order Autoregressive disturbances,

test $\quad H_0: \rho = 0 \quad$ vs. $\quad H_A: \rho > 0$.

—— normally a one-sided test.

1.) Asymptotic Test

C–O method;

$$\hat{\rho} = \frac{\sum_{t=2}^{T} e_t e_{t-1}}{\sum_{t=1}^{T} e_t^2} \quad \rightarrow \boxed{\text{consistent}}$$

Intuition;

regress e_t on e_{t-1} $\quad (t = 2, \ldots, T)$.

Then $\sqrt{T}(\hat{\rho} - \rho) \rightarrow N(0, 1-\rho^2)$. [by early Thm]

$\therefore$ Asymptotic Variance $= \dfrac{1-\rho^2}{T}$.

$$\Longrightarrow \quad \frac{\hat{\rho}}{\sqrt{\frac{1-\hat{\rho}^2}{T}}} \longrightarrow N(0,1)$$

under $H_0: \rho = 0$.

i.e. compare $\hat{\rho}$ to its standard error!

* This is easy, but no one does it;
need sample size > 30 to use it ["[illegible]

xerox

2.) Durbin-Watson test (popular)

This is an exact test
(as opposed to Asymptotic).

$$d = \frac{\sum_{t=2}^{T} (e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2}$$

Comparison to $\hat{\rho}$ in 1.) ;

$$\text{Note: } 2(1-\hat{\rho}) = 2\left[1 - \frac{\sum_{t=2}^{T} e_t e_{t-1}}{\sum_{t=1}^{T} e_t^2}\right]$$

$$= 2 - \frac{2\sum_{t=2}^{T} e_t e_{t-1}}{\sum_{t=1}^{T} e_t^2}$$

But expanding the numerator in d;

$$d = \frac{\sum_{t=2}^{T} e_t^2 + \sum_{t=1}^{T-1} e_t^2 - 2\sum_{t=2}^{T} e_t e_{t-1}}{\sum_{t=1}^{T} e_t^2}$$

Observe that <u>if</u> the first two sums were from 1 to T, then

$$d = \frac{2\sum_{t=1}^{T} e_t^2 - 2\sum_{t=2}^{T} e_t e_{t-1}}{\sum_{t=1}^{T} e_t^2} = 2 - \frac{2\sum_{t=2}^{T} e_t e_{t-1}}{\sum_{t=1}^{T} e_t^2} = \underline{\underline{2(1-\hat{\rho})}}$$

Thus $d \approx 2(1-\hat{\rho})$,

and if $\hat{\rho}$ is significantly different from 0, d will be different from 2.

$\therefore$ We'll reject $H_0: \rho = 0$ [no autocorrelation] if $d <$ some critical value, d^*_α. (significantly < 2)

Again ρ is rarely < 0, so we'll typically be concerned with d being significantly < 2.

Further Intuition:

Observe, $d \approx \dfrac{Var\{e_t - e_{t-1}\}}{Var\{e_t\}}$

If $\nexists$ no autocorrelation, then numerator $\approx 2 *$ denominator.

If $\exists$ ~~positive~~ autocorrelation,

$$Var\{e_t - e_{t-1}\} = Var\{e_t\} - 2\,Cov(e_t, e_{t-1}) + Var\{e_{t-1}\}$$

and $d \approx 1 - 2\dfrac{Cov\{e_t, e_{t-1}\}}{Var\{e_t\}} + 1 = 2\left[1 - \hat{\rho}(e_t, e_{t-1})\right]$

If $\exists$ perfect positive correlation, $\hat{\rho} = 1$,
and $d = 0$.

If $\exists$ perfect negative correlation, $\hat{\rho} = -1$,
and $d = 4$.

Thus the large sample limits of d are $\{0, 4\}$.

Comment: the distribution of d
is complicated and depends on the X's,
so the critical values depend on the X's.

i.e. Unlike a t-statistic,
which is compared to 2;
the critical values of d
depend on the model estimated.

For this reason, people have computed bounds
that depend only on T & K.

Consider the true critical point, d_α^*;

$P(d < d_\alpha^*) \geq \alpha$ under $H_0: \rho = 0$.

Then

$$(d_\alpha^*)_L \leq d_\alpha^* \leq (d_\alpha^*)_U$$

These are the <u>bounds</u> of this critical value, d_α^*.

$\exists$ Tables of $(d_\alpha^*)_L$ & $(d_\alpha^*)_U$ that, again, depend on T & K.

Reject H_0 if $d < (d_\alpha^*)_L$;
Accept H_0 if $d > (d_\alpha^*)_U$.

Inconclusive Region – between the bounds.

☆ Conservative procedure would Reject H_0 if d is below $(d_\alpha^*)_U$.

<u>Fact:</u> The inconclusive region is <u>smaller</u> if the # of degrees of freedom ($T-K$) is <u>bigger</u> [more obs., fewer regressors].

— & visa versa.

<u>Note:</u> Most Tables of these critical bounds apply to regressions with a <u>constant term</u>. [$\exists$ Tables for models with no constant.] Furthermore, K generally refers to the # of regressors in addition to the constant.

Comments :

① Autocorrelation is not a difficult problem to solve.
We don't lose much by using GLS (unless the sample size is ridiculously small), & we gain much.
— efficient estimates, appropriate tests.

——— $\exists$ too much literature about this problem.

② Most of these tests (including DW) presume First Order Autocorrelation, which does not always apply.

If not first order, use parametric tests [see Johnston].

ALTERNATIVE Error Structures

–[to First Order Autocorrelation]

(i) m^{th} Order Autoregressive Process

$$\epsilon_t = \rho_1 \epsilon_{t-1} + \rho_2 \epsilon_{t-2} + \cdots + \rho_m \epsilon_{t-m} + u_t$$

with u_t iid $N(0, \sigma^2)$

This is easy to work with.
The analogue of the C-O transformation works.

e.g. $m = 2$; $y_t = \alpha + \beta x_t + \epsilon_t$
where $\epsilon_t = \rho_1 \epsilon_{t-1} + \rho_2 \epsilon_{t-2} + u_t$

GLS transformation ;

$$(y_t - \rho_1 y_{t-1} - \rho_2 y_{t-2}) = \alpha(1 - \rho_1 - \rho_2) + \beta(x_t - \rho_1 x_{t-1} - \rho_2 x_{t-2}) + (\epsilon_t - \rho_1 \epsilon_{t-1} - \rho_2 \epsilon_{t-2})$$

for $t = 3, 4, \ldots, T$ [* New disturbance $= u_t$.]

(ii) n^{th} Order Moving Average Process

$$\epsilon_t = u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \cdots + \theta_n u_{t-n}$$

— Sort of a backwards representation of the n^{th} order autoregressive process.

Harder to handle ; later under time series analysis.

(iii) ARMA (m, n)

— Autoregressive Moving Average Process

$$\epsilon_t = \rho_1 \epsilon_{t-1} + \cdots + \rho_m \epsilon_{t-m} + u_t + \theta_1 u_{t-1} + \cdots + \theta_n u_{t-n}$$

— like (ii), harder to handle;
must use nonlinear L.S. procedures;
programs to find top of hill...
See Box & Jenkins; time series analysis.

(iv) Free Form Assumption

Hannan; Spectral Analysis.

– a way to run GLS w/o any "strict" assumptions on the disturbances.

– Covariance Stationary Analysis

Again; the covariance between any two terms the same distance apart are equal.
1-pd. apart covariances all the same.
2- " " " " " ".
⋮

Ω is a band matrix;
off-diagonal diagonals are constants.

a) $Cov(\epsilon_t, \epsilon_{t-K})$ is constant $\forall\ t$
b) $Cov(\epsilon_t, \epsilon_{t-K})$ ↓es as K ↑es.

(b) is the reason we need only a few bands, in any autoregressive process.

Note: The C-O transformation, V, has the diagonal and just one band.

Considerations

What does existence of autocorrelation imply?

1. Could be result of Specification Error.

   If this is the case (e.g. omitted variables), then the solution to Autocorrelation [GLS] will not solve the omitted variables problem!

2. Random disturbances are simply correlated for some reason; then conceivably the autoregressive disturbance could be negatively correlated as well as positively.

   Comment: in cross-section data, you'd better think of a good reason (that would depend on the ordering pattern)!