Regression with Correlated Errors

\[ y_t = \alpha + \beta x_t + e_t \]

• In some regression models, the errors are correlated
  – Pure Trend Models
  – Pure Seasonality Models

• In these models the errors can be correlated

• Classical and robust standard errors are not appropriate
Example: Stock Volume
Least-Squares Variance Formula

Recall for $\nu_t = x_t e_t$

$$\text{var}(\hat{\beta}) \sim \frac{\text{var} \left( \sum_{t=1}^{T} \nu_t \right)}{[T \text{var}(x_t)]^2}$$

When the $\nu$ are uncorrelated

$$\text{var} \left( \sum_{t=1}^{T} \nu_t \right) = \sum_{t=1}^{T} \text{var}(\nu_t) = T \text{var}(\nu_t)$$

$$\text{var}(\hat{\beta}) \sim \frac{\text{var}(\nu_t)}{T[\text{var}(x_t)]^2}$$
General Formula

Define

\[
f_T = \frac{\text{var} \left( \sum_{t=1}^{T} v_t \right)}{T \text{ var}(v_t)}
\]

When the \( v \) are uncorrelated \( f_T = 1 \), otherwise not.

Then

\[
\text{var}(\hat{\beta}) \sim \frac{\text{var}(x_t e_t)}{T \left[ \text{var}(x_t) \right]^2} f_T
\]
Adjustment Factor

• The asymptotic variance of least-squares is the conventional, multiplied by an adjustment factor for the serial correlation

\[
\text{var}(\hat{\beta}) \sim \frac{\text{var}(x_t e_t)}{T [\text{var}(x_t)]^2} f_T
\]
Autocovariance of $v$

- We want a useful formula for
  \[
  f_T = \frac{\text{var} \left( \sum_{t=1}^{T} v_t \right)}{T \text{var}(v_t)}
  \]

- Since $E(v_t) = 0$, then
  \[
  E(v_t^2) = \text{var}(v_t)
  \]
  \[
  E(v_t v_j) = \text{cov}(v_t v_j) = \gamma(t - j)
  \]

the autocovariance of $v_t$
Variance of sum of correlated $v$

$$\text{var}\left(\sum_{t=1}^{T} v_t\right) = E\left(\sum_{t=1}^{T} v_t\right)^2$$

$$= E\left(\sum_{t=1}^{T} v_t \sum_{j=1}^{T} v_j\right)$$

$$= \sum_{t=1}^{T} \sum_{j=1}^{T} E(v_t v_j)$$

$$= \sum_{t=1}^{T} \sum_{j=1}^{T} \gamma(t - j)$$
Adjustment Factor

\[ f_T = \frac{\text{var} \left( \sum_{t=1}^{T} v_t \right)}{T \text{var}(v_t)} = \frac{1}{T} \sum_{t=1}^{T} \sum_{j=1}^{T} \rho(t-j) \]

- Where the \( \rho(t-j) \) are the autocorrelations of \( v_t \)
This double sum is the sum of all the elements in the matrix

\[
\begin{bmatrix}
\rho(0) & \rho(1) & \rho(2) & \cdots & \rho(T-1) \\
\rho(1) & \rho(0) & \rho(1) & \cdots & \rho(T-2) \\
\rho(2) & \rho(1) & \rho(0) & \cdots & \rho(T-3) \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\rho(T-1) & \rho(T-2) & \rho(T-3) & \cdots & \rho(0)
\end{bmatrix}
\]

There are

- \( T \) of the \( \rho(0) \)
- \( 2(T-1) \) of the \( \rho(1) \)
- \( 2(T-2) \) of the \( \rho(2) \)
- \( \ldots \)

\[ T + \sum_{j=1}^{T-1} 2(T-j)\rho(j) \]
Adjustment Factor

- Dividing by $T$

\[ f_T = \frac{1}{T} \sum_{t=1}^{T} \sum_{j=1}^{T} \rho(t - j) \]

\[ = 1 + \sum_{j=1}^{T-1} 2 \left( \frac{T - j}{T} \right) \rho(j) \]

- If $T$ is large

\[ f_T \rightarrow 1 + 2 \sum_{j=1}^{\infty} \rho(j) = f \]
Summary: Least-Squares Variance

• When the errors are correlated

\[ \text{var}(\hat{\beta}) \sim \frac{\text{var}(x_t e_t)}{T \text{var}(x_t)^2} f \]

\[ f = 1 + 2 \sum_{j=1}^{\infty} \rho(j) \]

• The conventional formula is multiplied by an adjustment for autocorrelation
HAC Estimation

• Estimation of $f$
  – For variances and standard errors under autocorrelation

• Called heteroskedasticity and autocorrelation consistent (HAC) variance estimation

• Multiply conventional variance estimates by estimates of $f$
HAC Estimation

- The adjustment is

\[ f = 1 + 2 \sum_{j=1}^{\infty} \rho(j) \]

where \( \rho(j) \) are the autocorrelations of \( v_t = x_t e_t \)

- Estimate \( \rho(j) \) by sample autocorrelations using least-squares residuals

- But in a sample of length \( T \) we cannot estimate all autocorrelations well
Unweighted HAC Estimator

• For some *truncation parameter* \( m \),

\[
\hat{f} = 1 + 2 \sum_{j=1}^{m} \hat{\rho}(j)
\]

• Original proposal
  – L. Hansen, Hodrick (1978)
  – Hal White (1982)

• Deficiencies
  – This estimator is not smooth in the truncation parameter
  – The sample estimate can be negative
Lars Hansen

- Professor Lars Hansen, U Chicago
- Invented Generalized Method of Moments, the leading estimation method for applied econometrics
- Introduced unweighted HAC estimator for multi-step regression models
- Won 2013 Nobel Prize in economics
Example of Negative Estimate

• Take \( m=1 \)
• Then \( \hat{f} = 1 + 2\hat{\rho}(1) < 0 \)
  if estimated \( \rho(1) < -1/2 \)
Example: Liquor Sales

- Transform to growth rates
- Monthly change in log liquor sales
- Regress on Seasonal Dummies only to obtain seasonal pattern
Autocorrelation of Residual

- The first autocorrelation is less than $-1/2$
Weighted HAC Estimator

\[ \hat{f} = 1 + 2 \sum_{j=1}^{m} \left( \frac{m - j}{m} \right) \hat{\varrho}(j) \]

- Called Newey-West variance estimator
  - Whitney Newey, Ken West (1987)
- This weighted estimator is always positive
- Smoothly changes in truncation parameter \( m \)
Whitney Newey and Ken West

• Professor Whitney Newey, MIT
  – Leading econometric theorist
• Professor Ken West, Wisconsin
  – Macroeconomist & econometrician
  – Forecast evaluation and comparison
• Joint paper in 1987
  – Weighted HAC estimator
  – One of the most referenced papers in econometrics
Computation

• In STATA, replace `regress` command with `newey` command
  
  `.newey y x, lag(m)`

• You supply the truncation parameter “m”

• Similar to regression with robust standard errors

• These are identical
  
  `.newey y x, lag(0)`

  `.reg y x, r`
Example: Liquor Sales

```
.reg dy b12.m,r
```

Linear regression

|    | Coef.   | Robust Std. Err. | t   | P>|t| | [95% Conf. Interval] |
|----|---------|------------------|-----|------|----------------------|
| dy |         |                  |     |      |                      |
| m  | -.788371 | .0120765         | -65.28 | 0.000 | -.8121825 to -.7645595 |
| 1  | -.321870 | .0105478         | -30.52 | 0.000 | -.3426677 to -.3010733 |
| 2  | -.210318 | .0094619         | -22.23 | 0.000 | -.2289744 to -.1916619 |
| 3  | -.300291 | .010514          | -28.56 | 0.000 | -.3210222 to -.2795607 |
| 4  | -.225811 | .0100036         | -22.57 | 0.000 | -.245536 to -.2060876 |
| 5  | -.318535 | .0096047         | -33.16 | 0.000 | -.3374735 to -.2995981 |
| 6  | -.261882 | .0100737         | -26.00 | 0.000 | -.2817449 to -.2420198 |
| 7  | -.339259 | .0107775         | -31.48 | 0.000 | -.3605093 to -.3180088 |
| 8  | -.362447 | .0123023         | -29.46 | 0.000 | -.3867042 to -.3381907 |
| 9  | -.278295 | .010299          | -27.02 | 0.000 | -.2986023 to -.257989 |
| 10 | -.276187 | .0108553         | -25.44 | 0.000 | -.2975908 to -.2547835 |
| _cons | .3099733 | .0065735         | 47.16 | 0.000 | .2970122 to .3229343 |

Number of obs = 215
F( 11,  203) = 423.80
Prob > F = 0.0000
R-squared = 0.9613
Root MSE = 0.0347
With Newey-West standard errors

```
. newey dy b12.m, lag(12)
```

Regression with Newey-West standard errors

| dy |       Coef. | Std. Err. | t     | P>|t| | [95% Conf. Interval] |
|----|-------------|-----------|-------|-----|---------------------|
| m  | -0.788371   | 0.0149943 | -52.58| 0.000| -0.8179356, -0.7588064 |
| 1  | -0.3218705  | 0.0093479 | -34.43| 0.000| -0.3403018, -0.3034391 |
| 2  | -0.2103181  | 0.0100234 | -20.98| 0.000| -0.2300816, -0.1905547 |
| 3  | -0.3002915  | 0.0087418 | -34.35| 0.000| -0.3175278, -0.2830551 |
| 4  | -0.2258118  | 0.0128307 | -17.60| 0.000| -0.2511104, -0.2005132 |
| 5  | -0.3185358  | 0.0087245 | -36.51| 0.000| -0.335738, -0.3013336 |
| 6  | -0.2618824  | 0.0090442 | -28.96| 0.000| -0.279715, -0.2440498 |
| 7  | -0.3392591  | 0.0134996 | -25.13| 0.000| -0.3658765, -0.3126416 |
| 8  | -0.3624475  | 0.0075171 | -48.22| 0.000| -0.377269, -0.3476259 |
| 9  | -0.2782956  | 0.0116472 | -23.89| 0.000| -0.3012606, -0.2553307 |
| 10 | -0.2761872  | 0.0126533 | -21.83| 0.000| -0.3011359, -0.2512384 |
| _cons | 0.3099733  | 0.0066381 | 46.70 | 0.000| 0.2968848, 0.3230618 |
Truncation Parameter

- $m$ should be large when autocorrelation is large
- Sophistical data-dependent methods to pick $m$ have been developed, but are not in STATA
- Stock-Watson default (explanatory x’s)
  \[ m = 0.75T^{1/3} \]
- Trend/Seasonal default
  \[ m = 1.4T^{1/3} \]
Derivation of Defaults

• Due to Andrews (1991)
• The optimal \( m \) minimizes the mean-squared error of the estimate of \( f \)
• When \( \nu_t \) is an AR(1) with coefficient \( \rho \), Andrews found the optimal \( m \) is

\[
m = CT^{1/3}
\]

\[
C = \left( \frac{6 \rho^2}{(1 - \rho^2)^2} \right)^{1/3}
\]
Donald Andrews

- Professor Donald Andrews, Yale
- Leading econometric theorist
- Contributions to time-series
  - Optimal selection of truncation parameter
  - Tests for structural change
Default Values

\[ m = CT^{1/3} \]

\[ C = \left( \frac{6\rho^2}{(1-\rho^2)^2} \right)^{1/3} \]

- **Stock-Watson**
  - If both \( x_t \) and \( e_t \) are AR(1) with coef \( \frac{1}{2} \), then \( v_t = x_t e_t \) has AR(1) coefficient \( \rho = .25 \). Plug this in, and \( C = .75 \)

- **Trend-Seasonal**
  - If \( x_t \) is trend and/or seasonal and \( e_t \) are AR(1) with coef \( \frac{1}{2} \), then \( v_t = x_t e_t \) has AR(1) coefficient \( \rho = .5 \). Plug this in, and \( C = 1.4 \)
Liquor Sales again

. dis 1.4*e(N)^{(1/3)}
8.387017

. newey dy b12.m, lag(8)

Regression with Newey-West standard errors
Number of obs = 215
F( 11,  203) = 736.19
Prob > F = 0.0000

| dy | Coef. | Std. Err. | t | P>|t| | [95% Conf. Interval] |
|----|-------|-----------|---|------|----------------------|
| m  |       |           |   |      |                      |
| 1  | -.788371 | .0146673 | -53.75 | 0.000 | -.8172907 | -.7594513 |
| 2  | -.3218705 | .0089781 | -35.85 | 0.000 | -.3395727 | -.3041682 |
| 3  | -.2103181 | .0097191 | -21.64 | 0.000 | -.2294815 | -.1911548 |
| 4  | -.3002915 | .0097151 | -30.91 | 0.000 | -.319447 | -.281136 |
| 5  | -.2258118 | .0116748 | -19.34 | 0.000 | -.2488312 | -.2027924 |
| 6  | -.3185358 | .0089588 | -35.56 | 0.000 | -.3362001 | -.3008715 |
| 7  | -.2618824 | .00916 | -28.59 | 0.000 | -.2799433 | -.2438214 |
| 8  | -.3392591 | .0126319 | -26.86 | 0.000 | -.3641655 | -.3143526 |
| 9  | -.3624475 | .0091312 | -39.69 | 0.000 | -.3804516 | -.3444434 |
| 10 | -.2782956 | .0106888 | -26.04 | 0.000 | -.2993709 | -.2572204 |
| 11 | -.2761872 | .0126343 | -21.86 | 0.000 | -.3010984 | -.2512759 |
| _cons | .3099733 | .0065735 | 47.16 | 0.000 | .2970122 | .3229343 |
Example: Men’s Labor Force Participation Rate, Trend Model
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
Summary

• In one-step-ahead forecast regressions
• If the errors are serially uncorrelated
  – Use Robust standard errors
    • reg with r option
• If the errors are correlated
  – Use Newey-West standard errors
    • newey y x, lag(m)
  – In pure trend or seasonality models
    • Set m=1.4T^{1/3}
  – In dynamic regression
    • Set m=.75T^{1/3}
h-step-ahead forecasts

• In the AR(1) Model

\[ y_t = \alpha + \beta y_{t-1} + e_t \]

• The optimal h-step forecasting regression takes the form

\[ y_t = \alpha + \beta^h y_{t-h} + u_t \]
\[ u_t = e_t + \beta e_{t-1} + \beta^2 e_{t-2} + \cdots + \beta^{h-1} e_{t-h+1} \]

• The error \( u_t \) is a correlated MA(h-1)
  – Unless \( \beta=0 \)
h-step-ahead models

• In any h-step model

\[ y_t = \alpha + \beta y_{t-h} + u_t \]

the variable \( v_t = y_{t-h} e_t \) is generally serially correlated

• Generally MA(h-1)

• Correct adjustment term

\[ f = 1 + 2 \sum_{j=1}^{h-1} \rho(j) \]
Newey-West Standard Errors

• Standard errors can be estimated using the Newey-West method

• Truncation parameter set to forecast horizon
  – \( m = h \)

\[
\hat{f} = 1 + 2\sum_{j=1}^{h-1} \left( \frac{h-j}{h} \right) \hat{\rho}(j)
\]
Example: Unemployment Rate

- 12-month-ahead forecast with 4 AR lags
  - Robust standard errors:

```
.reg ur L(12/15).ur, r
```

Linear regression

| Coef. | Robust Std. Err. | t    | P>|t| | [95% Conf. Interval] |
|-------|------------------|------|------|---------------------|
| ur    |                  |      |      |                     |
| L12.  | 1.686434         | .2920485 | 5.77 | 0.000              | 1.113072 2.259795 |
| L13.  | -.0698989        | .3908098 | -0.18| 0.858              | -.837153 .6973552 |
| L14.  | -.5401552        | .3461042 | -1.56| 0.119              | -1.219641 .1393309 |
| L15.  | -.4100512        | .2538791 | -1.62| 0.107              | -.9084772 .0883747 |
| _cons | 1.94875          | .1705347 | 11.43| 0.000              | 1.613949 2.28355  |

Number of obs = 730
F(  4,    725) = 139.36
Prob > F = 0.0000
R-squared = 0.4955
Root MSE = 1.1088
Example: Unemployment Rate

- Newey-West standard errors:
- Standard errors on lag 13 and 14 decrease by half
- Standard error on constant more than doubles

```
.newey ur L(12/15).ur, lag(12)
```

```
Regression with Newey-West standard errors
maximum lag: 12

|      | Coef.  | Std. Err. |    t  | P>|t| | [95% Conf. Interval] |
|------|--------|-----------|-------|------|----------------------|
| ur   |        |           |       |      |                      |
| L12  | 1.686434 | .273372   | 6.17  | 0.000| 1.149738 - 2.223129  |
| L13  | -0.0698989 | .1564772  | -0.45 | 0.655| -0.3771014 - 0.2373036 |
| L14  | -0.5401552 | .1378278  | -3.92 | 0.000| -0.8107445 - 0.2695658 |
| L15  | -0.4100512 | .246517   | -1.66 | 0.097| -0.8940236 - 0.0739212 |
| _cons| 1.94875  | .4550687  | 4.28  | 0.000| 1.05534 - 2.842159   |

Number of obs = 730
F( 4, 725) = 21.00
Prob > F = 0.0000
```
newey and forecasting

- `predict` works after `newey` command, but not with `stdf` option
- `e(rmse)` does not work, only after `regress` or `reg`
  - rmse not computed or reported
- `newey` not appropriate for iterated forecasts
- Use `newey` to assess model and examine coefficients
- Use `reg` to compute out-of-sample forecast intervals
Summary

• In one-step-ahead forecast regressions
  – If the errors are serially uncorrelated, use `r` option
  – If the errors are correlated
    • Use `newey` for standard errors
      – In pure trend or seasonality models set $m=1.4T^{1/3}$
      – In dynamic regression set $m=.75T^{1/3}n$
    • Use `reg` and `predict sf, stdf` for forecast intervals, or iterated forecasts with `forecast`

• In h-step-ahead forecast regressions
  – Use `newey` with $m=h$ for standard errors
  – Use `reg` and `predict sf, stdf` for forecast intervals
Joint Tests

\[ y_t = \alpha + \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p} + e_t \]

- How do we assess if a subset of coefficients are jointly zero? Example: 3\textsuperscript{rd}+4\textsuperscript{th} lags

```
. reg gdp L(1/4).gdp,r
Linear regression
Number of obs = 247
F(  4,  242) =  8.85
Prob > F    =  0.0000
R-squared   =  0.1584
Root MSE    =  3.8132

|       | Coef.   | Robust Std. Err. | t   | P>|t|   | [95% Conf. Interval] |
|-------|---------|-----------------|-----|------|---------------------|
| gdp   |         |                 |     |      |                     |
| L1.   | .327656 | .076895         | 4.26| 0.000| .1761871            | .479125   |
| L2.   | .1466135| .0858808        | 1.71| 0.089| -.0225558           | .3157828  |
| L3.   | -.0980287| .0728951    | -1.34| 0.180| -.2416186           | .0455611  |
| L4.   | -.0889209| .0790354      | -1.13| 0.262| -.244606            | .0667641  |
| _cons | 2.378427| .4731312       | 5.03| 0.000| 1.446447            | 3.310408  |
```
Joint Hypothesis

• This is a joint test of
  \[ \beta_3 = 0 \]
  \[ \beta_4 = 0 \]

• This can be done with an “F test”

• In STATA, after `regress (reg)` or `newey` .
  `test L3.gdp L4.gdp`

• List variables whose coefficients are tested for zero.
Joint Tests

• “F test” named after R.A. Fisher
  – (1890-1992)
  – A founder of modern statistical theory

• Modern form known as a “Wald test”, named after Abraham Wald (1902-1950)
  – Early contributor to econometrics
F test computation

```
. test L3.gdp L4.gdp
( 1)  L3.gdp = 0
( 2)  L4.gdp = 0

    F(  2,  242) =  1.76
    Prob > F =  0.1747
```

- You need to list each variable separately
- STATA describes the hypothesis
- The value of “F” is the F-statistic
- “Prob>F” is the p-value
  - Small p-values cause rejection of hypothesis of zero coefficients
  - Conventionally, reject hypothesis if p-value < 0.05
Example: 2-step-ahead GDP AR(4)

```
.newey gdp L(2/5).gdp, lag(2)
```

Regression with Newey-West standard errors

| Coef.  | Std. Err. | t    | P>|t|  | [95% Conf. Interval] |
|--------|-----------|------|------|-----------------------|
| gdp    | 0.2410617 | 0.0768239 | 3.14 | 0.002                  | 0.0897296 - 0.3923938 |
| L2.    | -0.0368004 | 0.0703583 | -0.52 | 0.601                 | -0.1753962 - 0.1017954 |
| L3.    | -0.0910108 | 0.0791053 | -1.15 | 0.251                 | -0.2468369 - 0.0648152 |
| L4.    | -0.1128763 | 0.0687243 | -1.64 | 0.102                 | -0.2482533 - 0.0225006 |
| _cons  | 3.329426   | 0.5460059 | 6.10 | 0.000                 | 2.253873 - 4.404979   |

```
.test L3.gdp L4.gdp L5.gdp
```

( 1) L3.gdp = 0
( 2) L4.gdp = 0
( 3) L5.gdp = 0

F( 3, 241) = 1.65
Prob > F = 0.1793
Testing after Estimation

• The commands `predict` and `test` are applied to the most recently estimated model

• The command `test` uses the standard error method specified by the estimation command
  – `reg y x`: classical F test
  – `reg r x, r`: heteroskedasticity-robust F test
  – `newey y x, lag(m)`: correlation-robust F test
    • (The robust tests are actually Wald statistics)