Integration

- Orders of Integration Terminology
  - A series with a unit root (a random walk) is said to be **integrated of order one**, or ***I*(1)**
  - A stationary series without a trend is said to be **integrated of order 0**, or ***I*(0)**
  - An ***I*(1)** series is differenced once to be ***I*(0)**
  - In general, we say that a series is ***I*(d)** if its d'th difference is stationary.

Integrated of order d

- A series is I(d) if

$$(1-L)^d y_t = z_t$$

is stationary and without trend.

- Examples
  - I(0): $y_t = z_t$
  - I(1): $(1-L)y_t = z_t$
  - I(2): $(1-L)^2 y_t = z_t$
- Possible I(2) series are price levels and money supply

Fractional Integration

- **Advanced side note!**
- We said a series is I(d) if

$$(1-L)^d y_t = z_t$$

- We did not require d to be an integer
- We say that y is **fractionally integrated** if 0<d<1 or -1<d<0
- A fractionally integrated series is in between I(0) and I(1)
- Strong dependence, slow autocorrelation decay
- Popular model for asset return volatility.

Fractional Differencing

- The fractional differencing operator is an infinite series

$$
\begin{aligned}
(1-B)^d &= \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k \\
&= \sum_{k=0}^{\infty} \frac{\prod_{a=0}^{k-1}(d-a)\ (-B)^k}{k!} \\
&= 1 - dB + \frac{d(d-1)}{2}B^2 - \cdots .
\end{aligned}
$$

Co-Integration

- We say that two series are co-integrated if a linear combination has a lower level of integration
- If y and x are each I(1), yet $z=y-\theta x$ is I(0)
- Example: Term Structure
  - We saw before that T3 appears to have a unit root
  - But the spread T12-T3 was stationary
  - T3 and T12 are co-integrated!

Common Co-Integration Relations

- Interest Rates of different maturities
- Stock prices and dividends
  - (Campbell and Shiller)
- Aggregate consumption and income
  - (Campbell and Shiller)
- Aggregate output, consumption, and investment
  - King, Plosser, Stock and Watson

Cointegrating Equation

- We said that y and x are cointegrated if

$$z_t = y_t - \theta x_t$$

  is stationary
- This is called the cointegrating equation
- θ is the cointegrating coefficient
- In some cases, θ is known from theory
  - often θ=1

Great Ratios

- If the aggregate variables Y and X are proportional in the long run, then

$$Z_t = \frac{Y_t}{X_t}$$

  is stationary.
- Then $\log(Z_t) = \log(Y_t) - \log(X_t)$

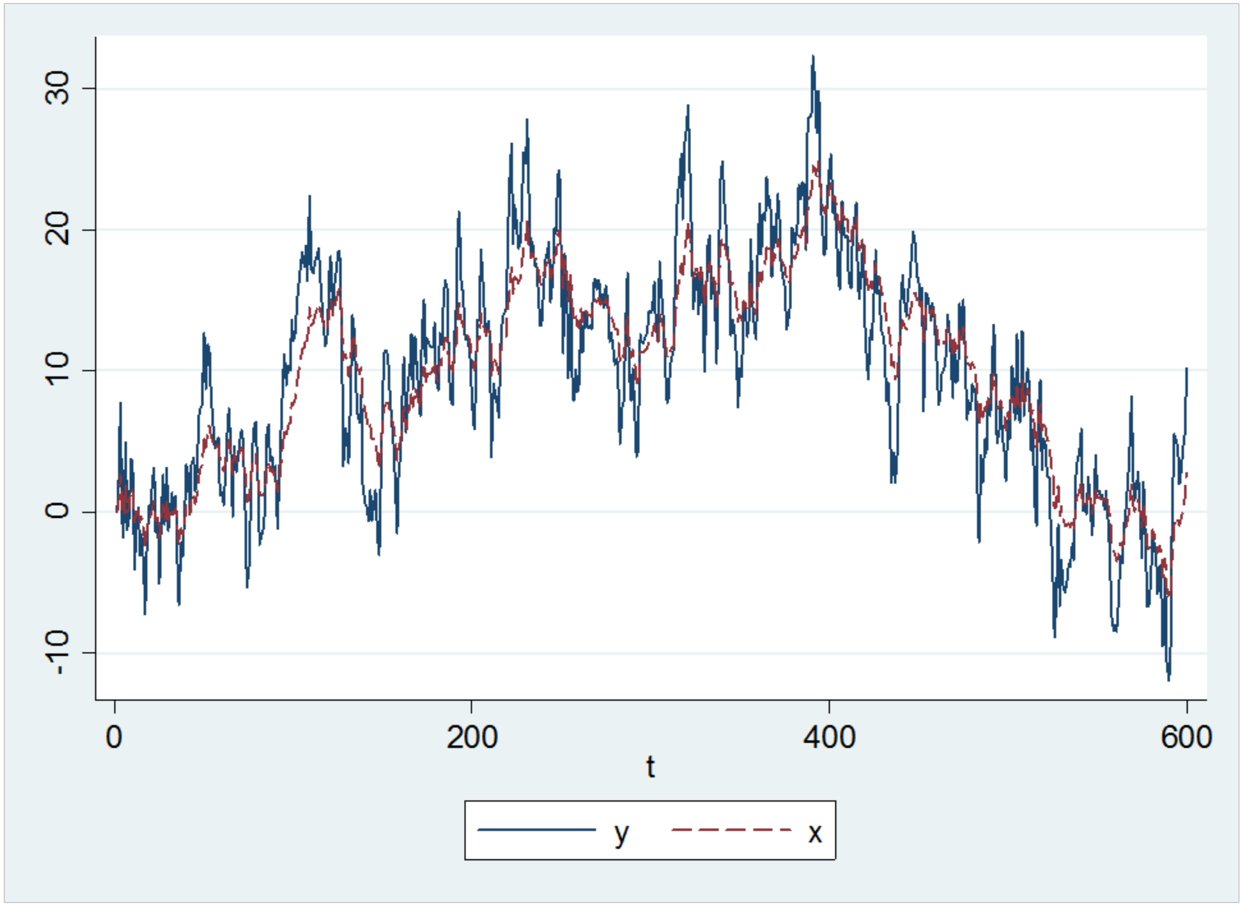
  and $z_t = y_t - x_t$

  where y=log(Y) and x=log(X)
- In this case, the logs y and x are cointegrated with coefficient 1.

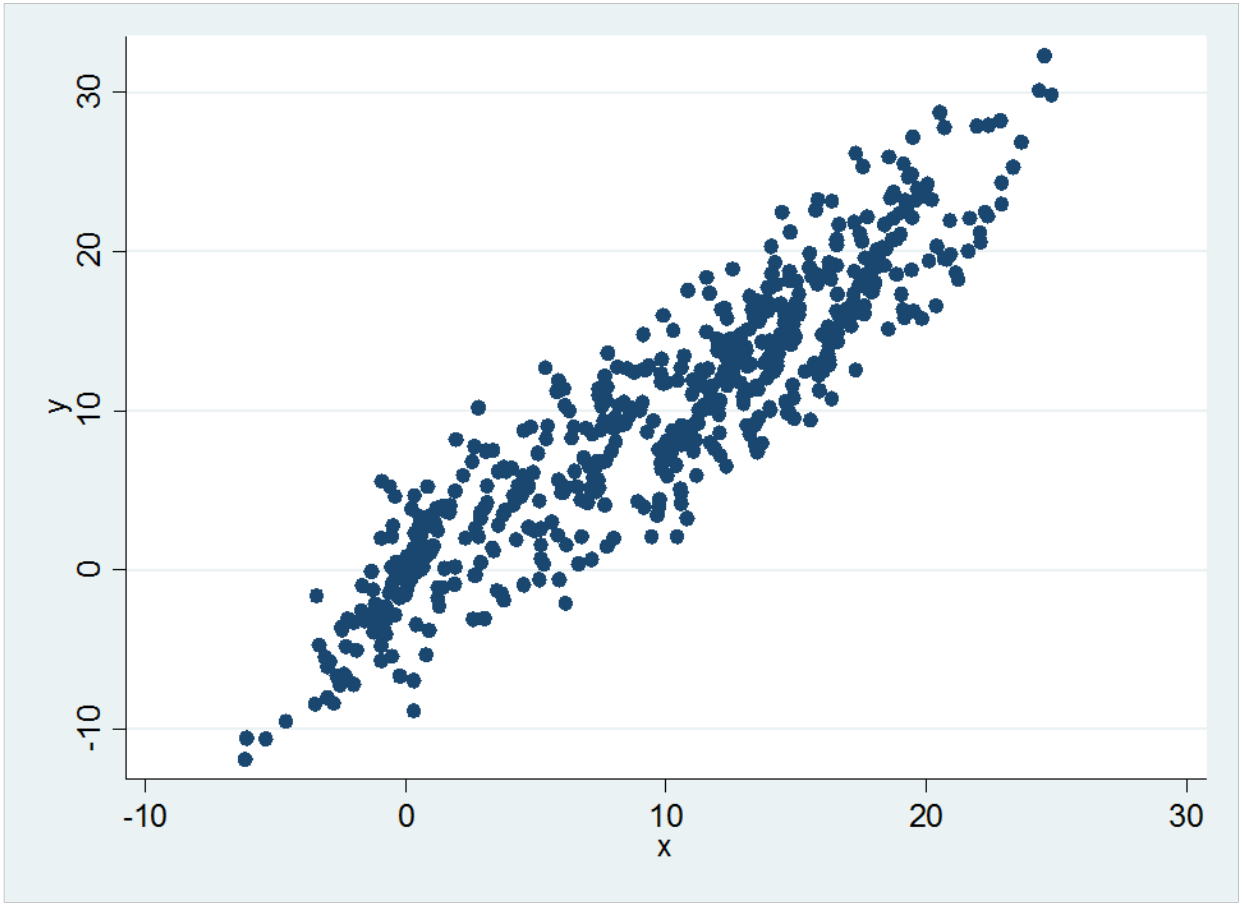
Equilibrium Error

- The difference $z_t = y_t - \theta x_t$ is sometimes called the equilibrium error, as it measures the deviation of y and x from the long-term cointegrating relationship

Simulated Example



Scatter plot
Variables stay close to cointegration line



Granger Representation Theory

- If y and x are I(1) and cointegrated, then the optimal regression for y takes the form

$$\Delta y_t = \mu + \gamma z_{t-1} + \alpha_1 \Delta y_{t-1} + \cdots + \alpha_p \Delta y_{t-p}$$

$$+ \beta_1 \Delta x_{t-1} + \cdots + \beta_q \Delta x_{t-q} + e_t$$

$$z_{t-1} = y_{t-1} - \theta x_{t-1}$$

- A dynamic regression in first differences, plus the error correction term z.

Answer to spurious regression

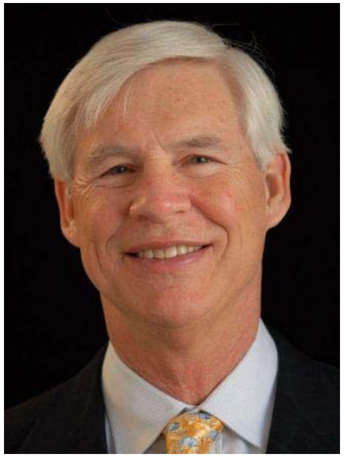
- The reaction to spurious regression was:
  - If the series are I(1), then do regressions in differences
- Cointegration says:
  - Add the error correction z!
- The difference is critical
  - The variable z measures if y is high or low relative to x
  - The error-correction coefficient γ pushes y back towards the cointegration relationship

Origin of Cointegration

- British econometricians
  - Davidson, Hendry, Srba and Yeo (1978)
  - Suggested $\ln(C_t)$-$\ln(Y_t)$ was a valuable predictor of consumption growth $\Delta\ln(C_t)$
  - This puzzled Clive Granger, as he knew that the variables were I(1), so should not be in a regression

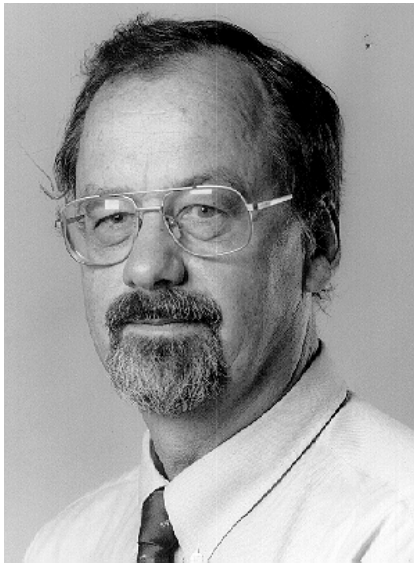
Theory of Cointegration

- This led Clive Granger to develop the theory of cointegration and the Granger Representation Theorem
- The most influential statement was a co-authored paper with Robert Engle (1987)
- Granger and Engle shared the Nobel Prize in economics in 2003



Cointegration Development

- Much of the statistical theory was developed by Peter Phillips and his students at Yale
- A multivariate statistical method was developed by the statistician Soren Johansen (U Copenhagen)
- Some jointly with the economist Katarina Juselius (Copenhagen)
- Their methods are programmed in STATA as VECM (vector error-correction models)



Example: Term Structure

- Regress change in 3-month T-bill on lagged spread, lagged changes in 3-month and 10-year
- Positive error correction coefficient
- Short rate increases when long rate exceeds short

```
. reg d.t3 L.spread120 L(1/12).d.t3 L(1/12).d.t120,r

Linear regression                                      Number of obs =     671
                                                       F( 25,   645) =    4.42
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.3032
                                                       Root MSE      =  .37838

------------------------------------------------------------------------------
             |               Robust
        D.t3 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   spread120 |
         L1. |   .0304828   .0186479     1.63   0.103    -.0061351    .0671007
```

Regression for Long Rate

- Long Rate decreases when long rate exceeds short

```
. reg d.t120 L.spread120 L(1/12).d.t3 L(1/12).d.t120,r

Linear regression                                      Number of obs =     671
                                                       F( 25,   645) =    5.04
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.2550
                                                       Root MSE      =  .24577
```

D.t120	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
spread120						
L1.	-.021608	.0114718	-1.88	0.060	-.0441345	.0009185

Unknown Cointegrating Coefficient

- If the cointegrating coefficient is unknown, it can be estimated
- Simplest estimator
  - Least squares of y on x
  - Consistent (Stock, 1987), but inefficient
  - Standard errors meaningless

```
. reg t3 t120

      Source |       SS       df       MS              Number of obs =     684
-------------+------------------------------           F(  1,   682) = 3240.27
       Model |  4663.61118     1  4663.61118           Prob > F      =  0.0000
    Residual |  981.579369   682  1.43926594           R-squared     =  0.8261
-------------+------------------------------           Adj R-squared =  0.8259
       Total |  5645.19055   683   8.2652863           Root MSE      =  1.1997

------------------------------------------------------------------------------
          t3 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        t120 |   .9708703   .0170557    56.92   0.000     .9373823    1.004358
       _cons |  -1.220776   .1175367   -10.39   0.000    -1.451554   -.9899991
------------------------------------------------------------------------------
```

Dynamic OLS

- Stock and Watson (1994) proposed a simple efficient estimator called dynamic OLS (DOLS)
- Regress y on x and leads and lags of Dx
- Use Newey-West standard errors
  - Lag $M=.75*T^{1/3}$
- **newey t3 t120 L(-12/12).d.t120, lag(6)**

Interest Rate Cointegration

```
. newey t3 t120 L(-12/12).d.t120, lag(6)

Regression with Newey-West standard errors          Number of obs  =       659
maximum lag: 6                                      F( 26,   632)  =     36.23
                                                    Prob > F       =    0.0000

------------------------------------------------------------------------------
             |             Newey-West
          t3 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        t120 |
         --. |   .9514564   .0377427    25.21   0.000     .8773402    1.025573
```

- The estimated cointegrating coefficient is 0.95
- The confidence interval contains our expected value of 1
- So in this case using the value 1 is recommended.

Estimated Cointegrating Coefficient

- Otherwise, the regression can use the estimated equilibrium error

$$z_{t-1} = y_{t-1} - \hat{\theta}x_{t-1}$$

Johansen VECM Method

- Alternatively, you can estimate the full VECM
- **vec t3 t120, trend(constant) lags(12)**
- This estimates a Vector Error Correction model with the variables T3 and T120, including a constant, and 12 lags of the variables
- This estimates equations for both variables, plus the cointegrating coefficient

Cointegrating Estimate

beta	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
_ce1						
t3	1	.	.	.	.	.
t120	-.9629871	.0747933	-12.88	0.000	-1.109579	-.8163949
_cons	1.198716	.	.	.	.	.

- The estimate is .96, similar to DOLS (.95)
- The DOLS method is simpler, but many econometricians prefer the VECM estimate.