Nonstationary Time Series and Cointegration

Lars A. Lochstoer

UCLA Anderson School of Management

Roadmap

1. Nonstationary time-series models and ARIMA
2. Unit Root Testing
3. Co-Integration
4. Vector Error-Correction Models (VECM)

ARIMA

Definition

A time series $\{y_t\}$ is said to be an $ARIMA(p, 1, q)$ if Δy_t is an $ARMA(p, q)$.

In this case, people call it **integrated of order 1** or $I(1)$.

- example:
  - y_t is non-stationary
  - $\Delta y_t = (1 - L)y_t = y_t - y_{t-1} = \eta_t$ is stationary
- in theory (and some applications), you might need to first difference a series multiple times: $ARIMA(p, d, q)$ or **integrated of order d**

Model with time trends and unit roots

- Consider the following model

$$y_t = c + \delta t + \phi_1 y_{t-1} + \varepsilon_t$$

- The model has a deterministic time trend: $c + \delta t$
- If $\phi_1 = 1$, it also has a unit root!
- Lesson: a model can (in theory) have two forms of non-stationarity.
  - trend stationarity and unit roots are not mutually exclusive
- many people think series like: U.S. log RGDP, CPI, and Consumption behave this way. See Cochrane (1988).

Dickey-Fuller (1979) test

- assume there is no deterministic time trend needed (e.g. exchange rates)
- suppose you consider one of the two models

$$p_t = \phi_1 p_{t-1} + \varepsilon_t \qquad \varepsilon_t \sim \mathrm{WN}(0, \sigma_\varepsilon^2)$$

$$p_t = \phi_0 + \phi_1 p_{t-1} + \varepsilon_t \qquad \varepsilon_t \sim \mathrm{WN}(0, \sigma_\varepsilon^2)$$

- we are interested in testing

$$H_0 : \phi_1 = 1 \qquad H_A : |\phi_1| < 1$$

- this is a one-sided test

Dickey-Fuller (1979) test

- We estimate the models by least squares.
- Let $\hat{\phi}_1$ denote the OLS estimator.
- t-ratio of least squares estimate:

$$t - ratio = \frac{\widehat{\phi}_1 - 1}{std(\widehat{\phi_1})}$$

Important:

- under the null hypothesis, the asymptotic sampling distributions of $\hat{\phi}_1$ and $t - ratio$ are **not** standard normal.
- under the null hypothesis, the data p_t is a random walk and is non-stationary....the standard central limit theorem does not apply.

Dickey-Fuller (1979) test

- **Important**: the sampling distributions depend on both the estimated model, the behavior under the null, and the alternative.
  - sampling distributions depend on whether you include a constant in the estimated regression.
  - sampling distributions depend on whether you include a drift in the null.
- How do you know whether to add a constant in the estimated regression?
- To choose the appropriate test and get the correct **critical values**, you need to understand what the trend properties of the data are under the alternative (next slide).

Dickey-Fuller (1979) Test: Summary

- Case 1:

$$\begin{aligned} Estimated &: \quad p_t = \phi_1 p_{t-1} + \varepsilon_t \\ Trueprocessundernull &: \quad p_t = p_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim \mathrm{WN}(0, \sigma_\varepsilon^2) \end{aligned}$$

- Case 2:

$$\begin{aligned} Estimated &: \quad p_t = \phi_0 + \phi_1 p_{t-1} + \varepsilon_t \\ Trueprocessundernull &: \quad p_t = p_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim \mathrm{WN}(0, \sigma_\varepsilon^2) \end{aligned}$$

- Case 3:

$$\begin{aligned} Estimated &: \quad p_t = \phi_0 + \phi_1 p_{t-1} + \varepsilon_t \\ Trueprocessundernull &: \quad p_t = \phi_0 + p_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim \mathrm{WN}(0, \sigma_\varepsilon^2) \end{aligned}$$

Testing with time trends

- Now, assume there is a time trend needed (e.g. log RGDP)
- we want to allow realistic trending behavior under the alternative hypothesis that the model is trend stationary.
- consider the following models

$$p_t = \delta t + \phi_1 p_{t-1} + \varepsilon_t \qquad \varepsilon_t \sim \mathrm{WN}(0, \sigma_\varepsilon^2)$$

$$p_t = c + \delta t + \phi_1 p_{t-1} + \varepsilon_t \qquad \varepsilon_t \sim \mathrm{WN}(0, \sigma_\varepsilon^2)$$

- we want to test the hypothesis:

$$H_0 : \phi_1 = 1 \qquad H_A : |\phi_1| < 1$$

Augmented Dickey-Fuller test: Said and Dickey (1984)

- Said and Dickey (1984) consider the following model with an $AR(p)$ structure:
$$p_t = c + \delta t + \phi_1 p_{t-1} + \sum_{i=1}^{p} \varphi_i \Delta p_{t-i} + \varepsilon_t$$
where ε_t denotes the error term.
- this model nests all of the possibilities we discussed above and more.
- the test allows for serially correlated data due to the AR(p) structure.
- they test the hypothesis
$$H_0 : \phi_1 = 1 \qquad \text{vs.} \qquad H_1 : \phi_1 < 1$$

Augmented Dickey-Fuller test: Said and Dickey (1984)

- t-ratio of least squares estimate:

$$\mathbf{t-ratio} = \frac{\widehat{\phi_1} - 1}{\mathrm{Std}(\widehat{\phi_1})}$$

- the sampling distribution of **t-ratio** is NOT standard.
- this implies:
  - critical values are not from a standard normal distribution.
  - critical values depend on whether you include:
    - a **constant** in the estimated regression!
    - a **time trend** in the estimated regression!

Right Specification?

- the asymptotic properties of the OLS estimate depend on whether you include in the estimated regression:
  - a constant
  - a time trend
- the asymptotic properties of the OLS estimate also depend on whether you include a drift term in the null hypothesis
- in the ADF test, you need to choose the lag order p. Ng and Perron (1995) and Ng and Perron (2001) discuss procedures for choosing p.

Right Specification?

- what is the correct specification?
  - if you're testing a specific hypothesis guided by theory, that would determine the correct specification
  - if not, look for a specification that makes sense under the null and the alternative hypothesis
    - ★ e.g. nominal interest rates: there is nothing in theory to suggest that nominal interest rates should have a deterministic time trend, but nominal interest rates do have a positive mean..so you should probably include a constant in estimated regression but no time trend.
    - ★ e.g. log U.S. RGDP: if the series does not have a unit root, the logarithm of real U.S. GDP does grow systematically over time. You probably want to include a time trend.

Long Run Relationships: Cointegration

- theory sometimes implies a long-run relation between different non-stationary variables
- Examples:
  - dividends and prices
  - purchasing power parity: price levels and exchange rates in different countries
  - permanent income theory: consumption, wealth and labor income
- **Co-Integration** was introduced by Engle and Granger (1987).

Stock Markets

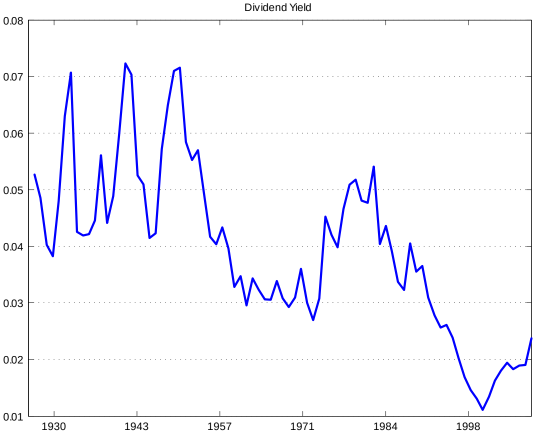
- dividends and stock prices are non-stationary
- however, we typically assume that there is a linear combination of (log) dividends and (log) prices

$$pd_t = p_t - d_t$$

that is stationary

  - In the absence of (long-lasting) bubbles, p_t and d_t should be co-integrated.
  - In fact, the evidence suggests that the price/dividend ratio is probably stationary, even though there is some evidence of structural breaks in the price/dividend ratio; see, Lettau and van Nieuwerburgh (2008).
  - If the price/dividend ratio is stationary, then a high price/dividend ratio should forecast either lower returns or higher dividend growth in the future (or a combination of both). [*see previous lectures*]

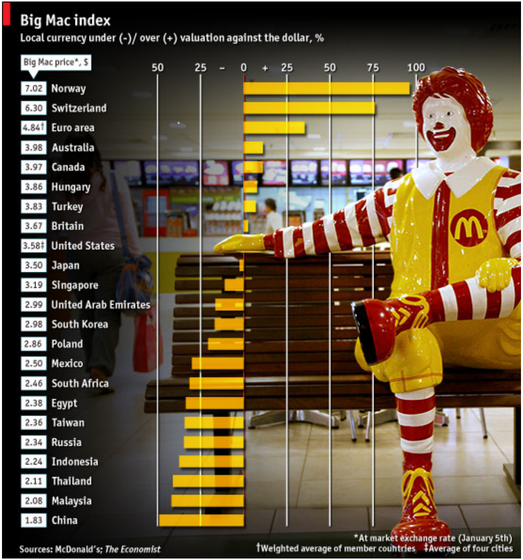
Dividend Yield



Dividend Yield on CRSP-VW (AMEX-NASDAQ-NYSE). Annual data. 1926-2008.

PPP

- purchasing-power parity (PPP): exchange rates should adjust to equalize the price of a basket of goods and services around the world.
- The Economist's Big Mac index: the "basket" is a burger, i.e. McDonald's Big Mac.


Big Mac index
Local currency under (-)/ over (+) valuation against the dollar, %
Big Mac price*, $
50
25
-
0
+
25
50
75
100
7.02 Norway
6.30 Switzerland
4.84† Euro area
3.98 Australia
3.97 Canada
3.86 Hungary
3.83 Turkey
3.67 Britain
3.58‡ United States
3.50 Japan
3.19 Singapore
2.99 United Arab Emirates
2.98 South Korea
2.86 Poland
2.50 Mexico
2.46 South Africa
2.38 Egypt
2.36 Taiwan
2.34 Russia
2.24 Indonesia
2.11 Thailand
2.08 Malaysia
1.83 China
Sources: McDonald's; The Economist
*At market exchange rate (January 5th)
†Weighted average of member countries ‡Average of four cities

PPP

- Here is an example on Jan 30, 2015:
  - the current SFR/dollar exchange rate is 0.86
  - the average SFR price of a burger in Switzerland was 6.50 SFR.
  - the average dollar price of a burger in Switzerland was \$7.54 > \$4.79 (the NY price of a Big Mac)
  - the PPP value of the SFR/dollar exchange rate is 1.36

$$P_t^{US,\$} = P_t^{Swi,SFR} \times S_t^{\$,SFR}$$

$$4.79\$ = 6.50 SFR \times 1.36^{-1}$$

Commodity Market Abritrage

- prices of goods and services offered in different locales can differ, but presumably they cannot wander off too far apart
- the percentage deviation from Big Mac PPP is given by:

$$y_t = \log P_t^{Swi,SFR} + \log S_t^{\$,SFR} - \log P_t^{US,\$}$$

  - arbitrageurs would step in to eliminate price differences
  - in international economics/finance this has given rise to the idea of purchasing power parity

- a weak version of PPP states that y_t should be stationary:

$$y_t = p_t^f + s_t - p_t$$

  where p_t denotes the log of the price level and p_t^f denotes the log of the foreign price level; s_t denotes the log of the exchange rate (dollars per unit of foreign currency)

PPP

- the foreign and domestic price level p^f and p^d are all integrated of order one ($I(1)$)
- so is the nominal exchange rate s_t
- but there is a linear combination of these 3 variables (e.g. the Swiss SFR price of a Big Mac, the NY dollar price of a Big Mac, and the SFR/Dollar exchange rate) that is stationary ($I(0)$)
- these variables (p_t^d, p_t^f, s_t) are co-integrated
- these variables share a common trend

Co-integration: Engle and Granger (1987)

Definition

Consider a vector of random variables $\{r_{1t}, r_{2t}, \ldots, r_{Nt}\}$.

Suppose all components are integrated of order d.

These variables are **co-integrated** if there exists a linear combination of $\mathbf{r}_t$ that is integrated of order $d - b$ where $b > 0$. This means that

$$\boldsymbol{\beta}' \mathbf{r}_t$$

is integrated of order $d - b$ for some vector $\boldsymbol{\beta}$. This is called the co-integrating vector.

- the number of cointegrating relations is b
- $\boldsymbol{\beta}$ is a $b \times N$ matrix.

Co-integration

- Typically, $d = 1$ and $b = 1$: we have a vector of unit-root variables and there is a linear combination of variables that is stationary
- the co-integrating vector is not unique. We can multiply the vector $\boldsymbol{\beta}$ by any non-zero scalar γ. For example,

$$\gamma\boldsymbol{\beta}$$

  is also a co-integrating vector.
- we normalize one of the elements in $\boldsymbol{\beta}$, i.e. set one of them equal to one.
- two variables integrated of different order cannot be co-integrated

PPP example

- PPP means that there exists a vector of $\boldsymbol{\beta} = (1, 1, -1)$ such that this linear combination
$$p_t^f + s_t - p_t^d$$
is stationary
- in stock markets, we assume the co-integrating vector is $\boldsymbol{\beta} = (1, -1)$ for (d_t, p_t), so that the log dividend yield is stationary

Shared Trend

Result

Suppose we have two-non stationary variables y_t and z_t. Then the two variables must have the same stochastic trend if they are co-integrated. This means that if

$$y_t = \mu_{yt} + \varepsilon_{yt}$$

$$z_t = \mu_{zt} + \varepsilon_{zt}$$

where μ_{it} is a random walk process and ε_{it} is stationary.

If y_t and z_t are co-integrated of order one, then they must have the same stochastic trend (see next slide).

Shared Trend

Proof.

We need to have non-zero β's such that $\beta_1 y_t + \beta_2 z_t$ is stationary:

$$\begin{aligned}\beta_1 y_t + \beta_2 z_t &= \beta_1 \left(\mu_{yt} + \varepsilon_{yt}\right) + \beta_2 \left(\mu_{zt} + \varepsilon_{zt}\right) \\ &= \left(\beta_1 \mu_{yt} + \beta_2 \mu_{zt}\right) + \left(\beta_1 \varepsilon_{zt} + \beta_2 \varepsilon_{yt}\right)\end{aligned}$$

Since the second term is stationary, the first term must satisfy:

$$\beta_1 \mu_{yt} + \beta_2 \mu_{zt} = 0$$

for all t.

This can only be true if the stochastic trends are the same (up to a scalar).

$$\mu_{yt} = -\frac{\beta_2}{\beta_1} \mu_{zt}$$

$\square$

Shared Trend

Result

Suppose we have three-non stationary variables y_t, z_t and w_t:

$$y_t = \mu_{y,t} + \varepsilon_{y,t}$$

$$z_t = \mu_{z,t} + \varepsilon_{z,t}$$

$$w_t = \mu_{w,t} + \varepsilon_{w,t}$$

where μ_{it} is a random walk process and ε_{it} is stationary. Suppose y_t, w_t and z_t are co-integrated of order one. Suppose that the trends satisfy:

$$\mu_{w,t} = \mu_{z,t} + \mu_{y,t}$$

Then the co-integrating vector is $\boldsymbol{\beta} = (1, 1, -1)$.

Co-integration and short-run dynamics

- if variables are co-integrated with vector $\boldsymbol{\beta} = (1, -\beta_{12})$ then the 'deviation from long run equilibrium' is

$$error_t = y_t - \beta_{12} z_t$$

  will influence the short-run dynamics of these variables

- PPP
  - if the price of a Brussels BIG Mac in dollars is too high relative to a NY Big Mac, then either the Brussels price in Euros of a Big Mac will decrease, the price in dollars of a NY Big Mac will increase or the Euro will depreciate...

$$error_t = p_t^f + s_t - p_t$$

PPP and VECMs

- consider a VAR for the p_t^f, p_t, s_t in first differences:

$$
\begin{aligned}
\Delta s_t &= \phi_{11}\Delta s_{t-1} + \phi_{12}\Delta p_{t-1} + \phi_{13}\Delta p_{t-1}^f + \varepsilon_{st} \\
\Delta p_t &= \phi_{21}\Delta s_{t-1} + \phi_{22}\Delta p_{t-1} + \phi_{23}\Delta p_{t-1}^f + \varepsilon_{pt} \\
\Delta p_t^f &= \phi_{31}\Delta s_{t-1} + \phi_{32}\Delta p_{t-1} + \phi_{33}\Delta p_{t-1}^f + \varepsilon_{pf,t}
\end{aligned}
$$

- this VAR is misspecified because it ignores the dynamics that push these variables back to the long-run equilibrium (PPP)

PPP

- here is a VAR for p_t^f, p_t, s_t with built-in error correction:

$$
\begin{aligned}
\Delta s_t &= \alpha_s \left(p_{t-1}^f + s_{t-1} - p_{t-1} \right) + \phi_{11} \Delta s_{t-1} + \phi_{12} \Delta p_{t-1} \\
&\quad + \phi_{13} \Delta p_{t-1}^f + \varepsilon_{st} \\
\Delta p_t &= \alpha_p \left(p_{t-1}^f + s_{t-1} - p_{t-1} \right) + \phi_{21} \Delta s_{t-1} + \phi_{22} \Delta p_{t-1} \\
&\quad + \phi_{23} \Delta p_{t-1}^f + \varepsilon_{pt} \\
\Delta p_t^f &= \alpha_{pf} \left(p_{t-1}^f + s_{t-1} - p_{t-1} \right) + \phi_{31} \Delta s_{t-1} + \phi_{32} \Delta p_{t-1} \\
&\quad + \phi_{33} \Delta p_{t-1}^f + \varepsilon_{pf,t}
\end{aligned}
$$

- this VAR incorporates the dynamics that push these variables back to the long-run equilibrium (PPP)

PPP

- the exchange rates, and price are ‘allowed to respond’ to the deviation from PPP
- the α’s are known as **speed of adjustment** parameters: if they are large, adjustment back to PPP happens faster!
- note that the α’s cannot all be zero

Error Correction Representation

Definition

A multivariate time series $\mathbf{r}_t$ has an **error correction representation** if

$$\boldsymbol{\Delta} r_t = \boldsymbol{\Phi}_0 + \underbrace{\boldsymbol{\Pi} r_{t-1}}_{errorcorrection} + \boldsymbol{\Phi}_1 \boldsymbol{\Delta} r_{t-1} + \ldots + \boldsymbol{\Phi}_p \boldsymbol{\Delta} r_{t-p} + \varepsilon_t, \varepsilon_t \sim \mathrm{WN}(0, \boldsymbol{\Sigma})$$

where $\boldsymbol{\Phi}_0$ is $N \times 1$ and $\boldsymbol{\Phi}_1, \ldots, \boldsymbol{\Phi}_p$ are $N \times N$.

The covariance matrix $\boldsymbol{\Sigma}$ is required to be positive definite.

- each row of $\boldsymbol{\Pi}$ is a co-integrating vector
- these rows are not linearly independent (if there is one co-integrating relation, then the rank of $\boldsymbol{\Pi}$ is one)
- if you leave out the error correction and these variables are co-integrated, the VAR in first differences is mis-specified

Error Correction Representation

Definition

A multivariate time series $\mathbf{r}_t$ has an **error correction representation** if

$$\Delta r_t = \mathbf{\Phi}_0 + \underbrace{\mathbf{\Pi} r_{t-1}}_{errorcorrection} + \mathbf{\Phi}_1 \Delta r_{t-1} + \ldots + \mathbf{\Phi}_p \Delta r_{t-p} + \varepsilon_t, \varepsilon_t \sim \text{WN}(0, \mathbf{\Sigma})$$

where $\mathbf{\Phi}_0$ is a N-dimensional vector, $\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_p$ are $N \times N$ matrices.

The variance covariance matrix is required to be positive definite.

- if $rank(\mathbf{\Pi}) = 0$, then Δr_t follows a VAR(p)
- if $rank(\mathbf{\Pi}) = N$, then $\mathbf{r}_t$ contains no unit roots
- if $rank(\mathbf{\Pi}) = b < N$, then $\mathbf{r}_t$ has b cointegrating relations

Error Correction Representation

Definition

$$\Delta r_t = \Phi_0 + \underbrace{\Pi r_{t-1}}_{errorcorrection} + \Phi_1 \Delta r_{t-1} + \ldots + \Phi_p \Delta r_{t-p} + \varepsilon_t, \varepsilon_t \sim \text{WN}\left(0, \Sigma\right)$$

The $N \times N$ matrix Π can be decomposed as

$$\Pi = \alpha \beta'$$

where α is a $N \times b$ matrix of **speed of adjustment** parameters and β' is a $b \times N$ matrix containing the cointegrating vectors.

- If there is cointegration, then Π contains fewer than N^2 free parameters. It is a restricted matrix.
- Cointegration models impose these restrictions.

From VAR(2) to VECM

- suppose $\mathbf{r}_t$ follows a VAR(2) as

$$\mathbf{r}_t = \mathbf{\Phi}_0^* + \mathbf{\Phi}_1^* \mathbf{r}_{t-1} + \mathbf{\Phi}_2^* \mathbf{r}_{t-2} + \boldsymbol{\varepsilon}_t \qquad \boldsymbol{\varepsilon}_t \sim \mathrm{WN}\left(0, \mathbf{\Sigma}\right)$$

- subtract $\mathbf{r}_{t-1}$ from both sides

$$\mathbf{r}_t - \mathbf{r}_{t-1} = \mathbf{\Phi}_0^* + \left(\mathbf{\Phi}_1^* - I_N\right) \mathbf{r}_{t-1} + \mathbf{\Phi}_2^* \mathbf{r}_{t-2} + \boldsymbol{\varepsilon}_t$$

- Let $\Delta \mathbf{r}_t = \mathbf{r}_t - \mathbf{r}_{t-1}$. Add/subtract $\Phi_2^* \mathbf{r}_{t-1}$ on the right

$$\begin{aligned} \Delta \mathbf{r}_t &= \mathbf{\Phi}_0^* + \left(\mathbf{\Phi}_2^* + \mathbf{\Phi}_1^* - I_N\right) \mathbf{r}_{t-1} - \mathbf{\Phi}_2^* \mathbf{r}_{t-1} + \mathbf{\Phi}_2^* \mathbf{r}_{t-2} + \boldsymbol{\varepsilon}_t \\ &= \mathbf{\Phi}_0 + \mathbf{\Pi} \mathbf{r}_{t-1} + \mathbf{\Phi}_1 \Delta \mathbf{r}_{t-1} + \boldsymbol{\varepsilon}_t \end{aligned}$$

where the relationship between the parameters is

$$\begin{aligned} \mathbf{\Phi}_0 &= \mathbf{\Phi}_0^* \\ \mathbf{\Pi} &= \mathbf{\Phi}_2^* + \mathbf{\Phi}_1^* - I_N \\ \mathbf{\Phi}_1 &= -\mathbf{\Phi}_2^* \end{aligned}$$

Error Correction Representation

- Engle and Granger propose a four-step procedure

  1. determine the order of integration
  2. estimate the long-run relationship
  3. estime the error-correction model
  4. assess model adequacy

- For details, see Hamilton (1994).

Order of Integration

- step 1: determine the order of integration (e.g. by running a battery of Dickey-Fuller tests)

Long-Run relation

- step 2: estimate the long-run relationship

$$y_t = \beta_0 + \beta_1 z_t + \varepsilon_t$$

  - if these variables are co-integrated, OLS is super-consistent (faster convergence); to determine if they actually are co-integrated, check the residuals
- do not use t-tests on slope coefficients (remember, badly scaled)

Check the Residuals

- take the estimated residuals from the regression of y_t on z_t and then run the following regression:

$$\Delta\widehat{\varepsilon}_t = a_1\widehat{\varepsilon}_{t-1} + v_t$$

- Null hypothesis of no cointegration: $H_0 : a_1 = 0$
- if we fail to reject the null that $a_1 = 0$, then we cannot reject the null of no co-integration (then the residuals contain a unit root)
- rejection of the null implies that the estimated residual series is stationary. The series appear cointegrated.

Estimation

- step 3: estimate the error-correction model: if the variables are integrated, then we can use the residuals to estimate the error-correction model

$$
\begin{aligned}
\Delta y_t &= \phi_{01} + \alpha_y \left(y_{t-1} - \hat{\beta} z_{t-1} \right) + \phi_{11} \Delta y_{t-1} + \phi_{12} \Delta z_{t-1} + \varepsilon_{yt} \\
\Delta z_t &= \phi_{02} + \alpha_z \left(y_{t-1} - \hat{\beta} z_{t-1} \right) + \phi_{21} \Delta y_{t-1} + \phi_{22} \Delta z_{t-1} + \varepsilon_{zt}
\end{aligned}
$$

  - we can use the estimated residual as a right hand side variable

$$
\begin{aligned}
\Delta y_t &= \phi_{01} + \alpha_y \widehat{\varepsilon}_{t-1} + \phi_{11} \Delta y_{t-1} + \phi_{12} \Delta z_{t-1} + \varepsilon_{yt} \\
\Delta z_t &= \phi_{02} + \alpha_z \widehat{\varepsilon}_{t-1} + \phi_{21} \Delta y_{t-1} + \phi_{22} \Delta z_{t-1} + \varepsilon_{zt}
\end{aligned}
$$

Assess Model Adequacy

- step 4: assess model adequacy by checking whether the residuals are white noise

References

Abadir, K. M. and J. R. Magnus (2005).
Matrix Algebra.
New York, NY: Cambridge University Press.

Christiano, L. J., M. Eichenbaum, and C. L. Evans (1999).
Monetary policy shocks: What have we learned and to what end?
In *Handbook of macroeconomics*. Elsevier.

Engle, R. F. and C. W. J. Granger (1987).
Co-integration and error correction: representation, estimation and testing.
Econometrica 55(2), 251–276.

Gatev, E., W. N. Goetzmann, and K. G. Rouwenhorst (2006).
Pairs trading: Performance of a relative-value arbitrage rule.
The Review of Financial Studies 19(3), 797–827.

Hamilton, J. D. (1994).
Time Series Analysis.
Princeton, NJ: Princeton University Press.

Lettau, M. and S. Van Nieuwerburgh (2008).
Reconciling the return predictability evidence.
The Review of Financial Studies 21(4), 1607–1652.

Vidyamurthy, G. (2004).
Pairs Trading: Quantitative Methods and Analysis.
Wiley Finance. Wiley Finance.