ECON 663: Econometrics II

Gabriel Rodriguez-Rondon
McGill University

Winter 2024

Outline

- Conditional Heteroskedastic Models
  - ARCH
  - GARCH
- ADL & Nonstationary Models
  - ADL Model
  - Cointegration
- Vector Autoregression (VAR) Models
  - VAR Models
  - Prediction
  - Causality
  - VECM
- Structural VAR (SVAR) Models
  - Identification problem
  - Short-run restrictions
  - Other identification methods

Conditional Heteroskedasticity Models

Based on: Hamilton (1994) Ch. 21; Verbeek (2004) Ch. 8; Martin et al. (2013) Ch. 20

See also: Enders (2015) Ch. 3

Introduction

So far, we have seen AR(p) models

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t \tag{1}$$

with the following properties

$$E[\epsilon_t] = 0 \qquad E[\epsilon_t \epsilon_s] = \begin{cases} \sigma^2 & \text{when } s = t \\ 0 & \text{when } s \neq t \end{cases}$$

and discussed conditions for these processes to be stationary (i.e., all roots outside the unit circle).

We know that when the process is stable $E[y_t] = \mu \ \forall t$ but that

$$E_t[y_t] = E[y_t|\mathcal{I}_t] = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p}$$

Similarly, although the unconditional variance $E[\epsilon_t^2] = \sigma^2$, we may be interested in the conditional variance $E_t[\epsilon_t^2] = \sigma_t^2$.

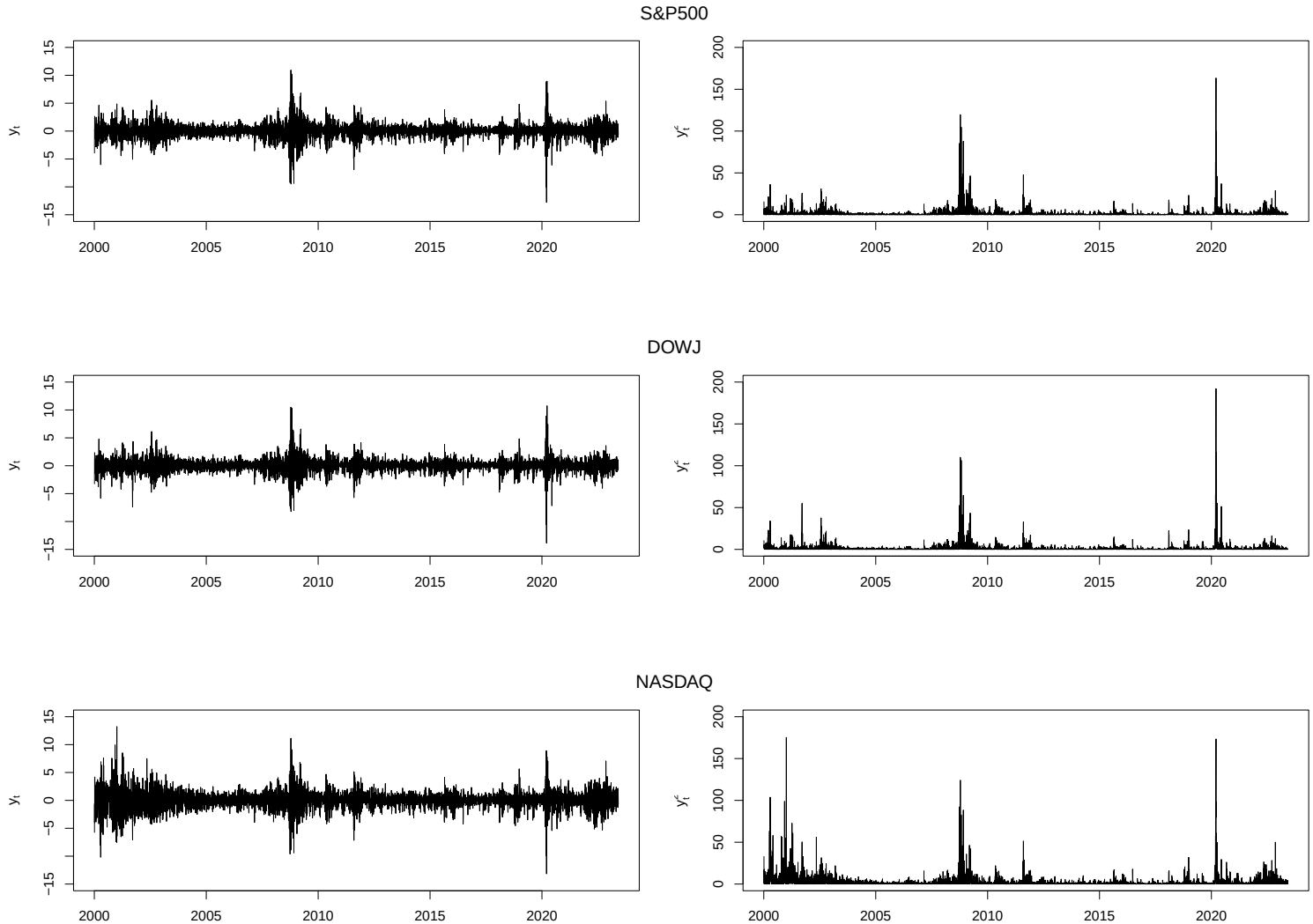
Introduction

Why should we care about conditional heteroskedasticity models?:

- Financial time series show evidence of volatility clustering
  - ▷ At high frequencies (e.g., minute-data, daily, weekly)
  - ▷ Big (small) shocks are followed by big (small) shocks in returns
- Relax assumption that variance is constant when modeling y_t or y_t^2
  - ▷ Volatility clustering gives rise to autocorrelation in y_t^2
- Conditional normality
  - ▷ Unconditional distribution will be Leptokurtic (fat tails)
  - ▷ Conditional distribution will be normal
- Better forecasts of ϵ_t and y_t^2

Introduction

Figure: Returns and squared returns of 3 stock indices



ARCH

We can describe the ϵ_t^2 using the following AR(p) model

$$\epsilon_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_q \epsilon_{t-q}^2 + \nu_t \tag{2}$$

where ν_t is another white noise process such that

$$E[\nu_t] = 0 \qquad E[\nu_t \nu_s] = \begin{cases} \lambda^2 & \text{when } s = t \\ 0 & \text{when } s \neq t \end{cases}$$

so that the conditional variance is given by

$$E_t[\epsilon_t^2] = \omega + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_q \epsilon_{t-q}^2 \tag{3}$$

A white noise process ϵ_t that is explained by the model given in (2) is described as an Autoregressive Conditional Heteroskedasticity (ARCH) model.

This class of models was first introduced by Engle (1982).

ARCH

We can also express the ARCH(q) model as

$$\epsilon_t = \sqrt{h_t} \cdot \nu_t \tag{4}$$

where instead

$$E[\nu_t] = 0 \qquad E[\nu_t \nu_s] = \begin{cases} 1 & \text{when } s = t \\ 0 & \text{when } s \neq t \end{cases}$$

and

$$h_t = \omega + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_q \epsilon_{t-q}^2 \tag{5}$$

This specification will allow us to estimate all parameters of the model for y_t and the parameters for h_t simultaneously.

ARCH

It can also be verified that the unconditional moments of ϵ_t are the same in both cases, regardless of whether we use (4) or (2)

That is,

$$E[\epsilon_t] = 0$$
$$E[\epsilon_t^2] = \frac{\omega}{(1 - \sum_{j=1}^{\infty} \alpha_j)} = \sigma^2$$

Unlike regular AR(p) models, restrictions on ω and α_j are needed. Specifically, since ϵ_t^2 and $E[\epsilon_t^2]$ must be positive, these parameters must all be positive.

ARCH: Maximum Likelihood Estimation

Suppose we want to estimate the following model by MLE

$$y_t = \boldsymbol{x}_t'\beta + \epsilon_t \tag{6}$$

$$\epsilon_t = \sqrt{h_t} \cdot \nu_t \tag{7}$$

where $\boldsymbol{x_t}$ is a vector of predetermined variables which may include lags of y_t and where ϵ_t^2 follows a ARCH(q). Assuming $\nu_t \sim i.i.d.\mathcal{N}(0,1)$, the conditional likelihood function is given by

$$f(y_t|\boldsymbol{x}_t, \mathcal{I}_{t-1}) = \frac{1}{\sqrt{2\pi h_t}} \exp\left(-\frac{1}{2}\frac{(y_t - \boldsymbol{x}_t\beta)^2}{h_t}\right) \tag{8}$$

where

$$h_t = \omega + \alpha_1(y_{t-1} - \boldsymbol{x}_{t-1}'\beta)^2 + \cdots + \alpha_1(y_{t-q} - \boldsymbol{x}_{t-q}'\beta)^2 \tag{9}$$

ARCH: Maximum Likelihood Estimation

Let $\theta = (\beta', \omega, \alpha_1, \ldots, \alpha_q)$. The conditional log likelihood to be maximized is given by

$$\mathcal{L}(\theta) = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\log(h_t) - \frac{1}{2}\sum_{t=1}^{T}\frac{(y_t - \boldsymbol{x}_t'\beta)^2}{h_t} \tag{10}$$

We can also work with non-Gaussian distributions.

For example, if we assume $\nu_t \sim i.i.d.t_\delta$, where δ is the degree of freedom parameter to be estimated and included in θ, then

$$\begin{aligned}\mathcal{L}(\theta) = T\log\left\{\frac{\Gamma[(\delta+1)/2]}{\pi^{1/2}\Gamma(\delta/2)}(\delta-2)^{-1/2}\right\} - \frac{1}{2}\sum_{t=1}^{T}\log(h_t) \\ -\frac{(\delta+1)}{2}\sum_{t=1}^{T}\left[1 + \frac{(y_t - \boldsymbol{x}_t'\beta)^2}{h_t(\delta-2)}\right]\end{aligned} \tag{11}$$

Alternatively, we can also use QMLE. See Hamilton (1994) Ch. 21 for a discussion of analytical or numerical solutions.

ARCH: Testing

So far, we are assuming

$$\epsilon \sim i.i.d.\mathcal{N}(0, \omega + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_q \epsilon_{t-q}^2)$$

If $\alpha_j = 0 \ \forall j$ then $\epsilon \sim i.i.d.\mathcal{N}(0, \omega)$ (i.e., no ARCH and the variance is constant). Hence, we can test for ARCH through the following hypothesis

$$H_0 : \alpha_1 = \cdots = \alpha_q = 0 \qquad \text{vs.} \qquad H_a : \alpha_j \neq 0 \text{ for some } j \tag{12}$$

This can be tested using a LR, Wald, or LM test. LM is commonly used as it only requires estimating (6) without ARCH.

ARCH: Testing

Although the conventional LM test statistic can be computed, Engle (1982) shows that in this case, an equivalent form of the LM statistic is given by

$$LM = T \cdot R^2 \tag{13}$$

Here, T is the sample size and R^2 is the coefficient of determination from the following regression

$$\hat{\epsilon}_t^2 = \omega + \alpha_1 \hat{\epsilon}_{t-1}^2 + \cdots + \alpha_q \hat{\epsilon}_{t-q}^2 + e_t \tag{14}$$

where $\hat{\epsilon}_t^2$ are the estimated residuals when estimating the model under the null.

Under the null,

$$LM \sim \chi^2_{(q)}$$

ARCH: Forecasting

Consider the following ARCH(1) model

$$\epsilon_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 \tag{15}$$

obtaining out-of-sample forecasts for the conditional variance from this ARCH(1) can be done as in previous chapters by taking the expectation conditional on information available at time t. That is,

$$E[\epsilon_{t+1}^2] = \epsilon_{t+1|t}^2 = \omega + \alpha_1 \epsilon_t^2$$

at $t = 2$

$$\begin{aligned} \epsilon_{t+2|t}^2 &= \omega + \alpha_1 \epsilon_{t+1|t}^2 \\ &= \omega + \alpha_1 \omega + \alpha_1^2 \epsilon_t^2 \end{aligned}$$

at $t = 3$

$$\begin{aligned} \epsilon_{t+3|t}^2 &= \omega + \alpha_1 \epsilon_{t+2|t}^2 \\ &= \omega + \alpha_1 \omega + \alpha_1^2 \omega + \alpha_1^3 \epsilon_t^2 \end{aligned}$$

or in general

$$\epsilon_{t+h|t}^2 = \omega \sum_{j=0}^{h-1} \alpha_1^j + \alpha_1^h \epsilon_t^2 \tag{16}$$

GARCH

If the process we want to model has long memory, we need to estimate an ARCH(q) with many lags q.

To circumvent this, we can use the Generalized ARCH (GARCH) model introduced by Bollerslev (1986). A GARCH(p,q) model is given by

$$\epsilon_t = \sqrt{h_t}\nu_t \tag{17}$$

$$h_t = \omega + \sum_{j=1}^{q} \alpha_j \epsilon_{t-j}^2 + \sum_{j=1}^{p} \beta_j h_{t-j} \tag{18}$$

Here, it is still the case that $E[\epsilon_t] = 0$, but now

$$E[\epsilon_t^2] = \frac{\omega}{\left(1 - \sum_{j=1}^{q} \alpha_j - \sum_{j=1}^{p} \beta_j\right)}$$

and so, as before, restrictions are still needed on α_j and β_j for this value to always be positive.

GARCH

In many cases, a GARCH(1,1) is appropriate. This is because, starting with a GARCH(1, 1) model

$$
\begin{aligned}
h_t &= \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 h_{t-1} \\
\beta(L) h_t &= \omega + \alpha_1 \epsilon_{t-1}^2 \\
h_t &= \beta(L)^{-1} \omega + \beta(L)^{-1} \alpha_1 \epsilon_{t-1}^2 \\
h_t &= \frac{\omega}{1-\beta_1} + \alpha_1 \sum_{i=0}^{\infty} \beta_1^i \epsilon_{t-1-i}^2
\end{aligned}
$$

we can obtain an ARCH(∞).

Note that the last line follows because, as seen before in Ch.8 of Verbeek (2004), the inverse of a finite order lag polynomial is an infinite order lag polynomial.

GARCH

Example from Verbeek (2004): US and Deutsche Mark Exchange rate from January 1980 to 21 May 1987

Table 8.6 GARCH estimates for change in log exchange rate US$/DM

	ARCH(6)	GARCH(1,1)		EGARCH
constant	0.228	0.016		−0.185
	(0.023)	(0.005)		(0.023)
ε_{t-1}^2	0.092	0.110	$\lvert\varepsilon_{t-1}\rvert/\sigma_{t-1}$	0.215
	(0.026)	(0.016)		(0.027)
ε_{t-2}^2	0.081	–		
	(0.025)			
ε_{t-3}^2	0.123	–		
	(0.028)			
ε_{t-4}^2	0.138	–		
	(0.033)			
ε_{t-5}^2	0.122	–		
	(0.028)			
ε_{t-6}^2	0.101	–		
	(0.027)			
σ_{t-1}^2	–	0.868	$\log(\sigma_{t-1}^2)$	0.968
		(0.018)		(0.009)
			$\varepsilon_{t-1}/\sigma_{t-1}$	−0.017
				(0.013)

GARCH: Testing

Given that it is more parsimonious, GARCH models are typically preferred. However, we can test for GARCH effects.

The same LM test as before for ARCH(q) is sometimes used with a large q as evidence for a GARCH(1,1) or GARCH(2,2).

Alternatively, note that if we have a GARCH(0,q) it is equivalent to an ARCH(q) model. Hence we can test for GARCH effects by testing the following hypothesis.

$$H_0 : \beta_1 = \cdots = \beta_p = 0 \qquad \text{vs.} \qquad H_a : \beta_j \neq 0 \text{ for some } j \tag{19}$$

This can be tested using a LR or Wald test.

GARCH

There are many other extensions of the basic ARCH model of Engle (1982). Some important extensions include:

- Exponential GARCH (EGARCH) of Nelson (1991) is used to consider asymmetric distributions

$$\log(h_t) = \omega + \alpha\left(\nu_{t-1} - E[\nu_{t-1}]\right) + \gamma\nu_{t-1} + \beta\log(h_{t-1}) \tag{20}$$

- Integrated GARCH (IGARCH) of Engle and Bollerslev (1986) is used for processes with unit-roots $\alpha + \beta = 1$ (so that $\beta = 1 - \alpha$)

$$h_t = \omega + \alpha\epsilon_{t-1}^2 + (1-\alpha)h_{t-1} \tag{21}$$

- Markov switching GARCH (MS-GARCH) to consider time-varying parameters (see Hamilton and Susmel (1994) and Gray (1996) among others)

$$h_t = \omega_{s_t} + \alpha_{s_t}\epsilon_{t-1}^2 + \beta_{s_t}h_{t-1} \tag{22}$$

See Bollerslev (2009) for a more complete list of the wide family of GARCH models.

Prediction

In practice, we will likely need compare the prediction of various models and try to determine which model provides better predictions.

Ways to compare forecasts, at different horizons h, include

- Comparing MSE
- Diebold Mariano (DM) Test (see Diebold and Mariano (2002))
  - Test for statistical difference in MSE
  - Easily applied to other loss functions other than MSE
- Model Confidence Set (MCS) Test (see Hansen et al. (2011))
  - Extension of DM test used to determine a set of optimal models

GARCH: Estimation & Forecasting using R

Figure: R output - MSFE of competing GARCH models

```
### Compute MSFE ------------------------------------------------------

# MSE of ARCH(1)
mean((SP500_yf^2 - SP500_arch1_fcast_fit@forecast$sigma^2)^2)
```

```
## [1] 1.679659
```

```
# MSE of ARCH(6)
mean((SP500_yf^2 - SP500_arch6_fcast_fit@forecast$sigma^2)^2)
```

```
## [1] 1.145585
```

```
# MSE of GARCH(1,1)
mean((SP500_yf^2 - SP500_garch_fcast_fit@forecast$sigma^2)^2)
```

```
## [1] 0.9473519
```

```
# MSE of EGARCH(1,1)
mean((SP500_yf^2- SP500_egarch_fcast_fit@forecast$sigma^2)^2)
```

```
## [1] 0.9539608
```

See files available on *MyCourses*.

Autoregressive Distributed Lag Models & Models with Nonstationary Variables

Based on: Verbeek (2004) Ch. 9; Enders (2015)

ADL Models

Consider the following model:

$$Y_t = \delta + \theta Y_{t-1} + \phi_0 X_t + \phi_1 X_{t-1} + \epsilon_t \tag{23}$$

where both Y_t and X_t are stationary variables. Here, we can see that the effect of X_t are distributed across several periods.

These models are useful for

- Describing dynamic effects of a change in X_t on current and future values of Y_t
- Determining the contemporaneous impact of one variable X_t (e.g., policy variable) on another variable of interest Y_t
- Determining long-run impact/equilibrium relationship between these variables

ADL Models

From eq. (23), assuming $|\theta| < 1$, we can get

- Impact multiplier

$$\frac{\partial Y_t}{\partial X_t} = \phi_0$$

- Long-run multiplier

$$\frac{\partial Y_{t+\infty}}{\partial X_t} = \frac{\phi_0 + \phi_1}{1 - \theta}$$

- Long-run equilibrium relation between Y and X

$$\begin{aligned} E[Y_t] &= \delta + \theta E[Y_t] + \phi_0 E[X_t] + \phi_1 E[X_t] \\ &= \frac{\delta}{1 - \theta} + \frac{\phi_0 + \phi_1}{1 - \theta} E[X_t] \end{aligned}$$

where stationary of Y_t and X_t is also used in the last point (i.e., $E[Y_t] = E[Y_{t-1}]$ and $E[X_t] = E[X_{t-1}]$).

ADL Models

We can also get an equivalent representation of the model given by (23) by subtracting Y_{t-1} from both sides

$$Y_t - Y_{t-1} = \delta + \theta Y_{t-1} + \phi_0 X_t + \phi_1 X_{t-1} + \epsilon_t - Y_{t-1}$$

$$\Delta Y_t = \phi_0 \Delta X_t - (1-\theta)[Y_{t-1} - \frac{\delta}{(1-\theta)} - \frac{(\phi_0+\phi_1)}{(1-\theta)} X_{t-1}] + \epsilon_t$$

or simply

$$\Delta Y_t = \phi_0 \Delta X_t - (1-\theta)[Y_{t-1} - \alpha - \beta X_{t-1}] + \epsilon_t \tag{24}$$

where $\alpha = \frac{\delta}{(1-\theta)}$ and $\beta = \frac{(\phi_0+\phi_1)}{(1-\theta)}$.

This is an example of an error-correction model.

ADL Models

$$\Delta Y_t = \phi_0 \Delta X_t - (1-\theta)[Y_{t-1} - \alpha - \beta X_{t-1}] + \epsilon_t$$

Here, the term $[Y_{t-1} - \alpha - \beta X_{t-1}]$ measures deviations from the long-run equilibrium (i.e., equilibrium error).

- If the equilibrium error term is positive, then $-(1-\theta)$ adjusts ΔY_t downwards
- If the equilibrium error term is negative, then $-(1-\theta)$ adjusts ΔY_t upwards

The stability condition $|\theta| < 1$ ensures that $(1 - \theta > 0)$. And so changes in Y_t are explained by current changes in X_t plus this error-correction term.

ADL Models

We can easily generalize the ADL model given in eq. (23) to include more lags

$$\theta(L)Y_t = \delta + \phi(L)X_t + \epsilon_t \tag{25}$$

where

$$\theta(L) = 1 - \theta_1 L - \cdots - \theta_p L^p$$
$$\phi(L) = \phi_0 + \phi_1 L + \cdots + \phi_q L^q$$

are lag polynomials but the constant term in $\phi(L)$ is not restricted to 1. Assuming we have a stationary process, $\theta(L)$ is invertible and so we can get

$$Y_t = \theta^{-1}(L)\delta + \theta^{-1}(L)\phi(L)X_t + \theta^{-1}(L)\epsilon_t$$

Now the coefficients of the lag polynomial $\theta^{-1}(L)\phi(L)$ determine the dynamic affects of X_t on Y_t and the long-run effect is now given by

$$\theta^{-1}(L)\phi(L) = \frac{\phi_0 + \phi_1 + \cdots + \phi_q}{1 - \theta_1 - \cdots - \theta_p}$$

Spurious Regression

Suppose now that Y_t and X_t are non-stationary variables, generate by independent random walks, and we attempt to estimate the following regression

$$Y_t = \alpha + \beta X_t + \epsilon_t$$

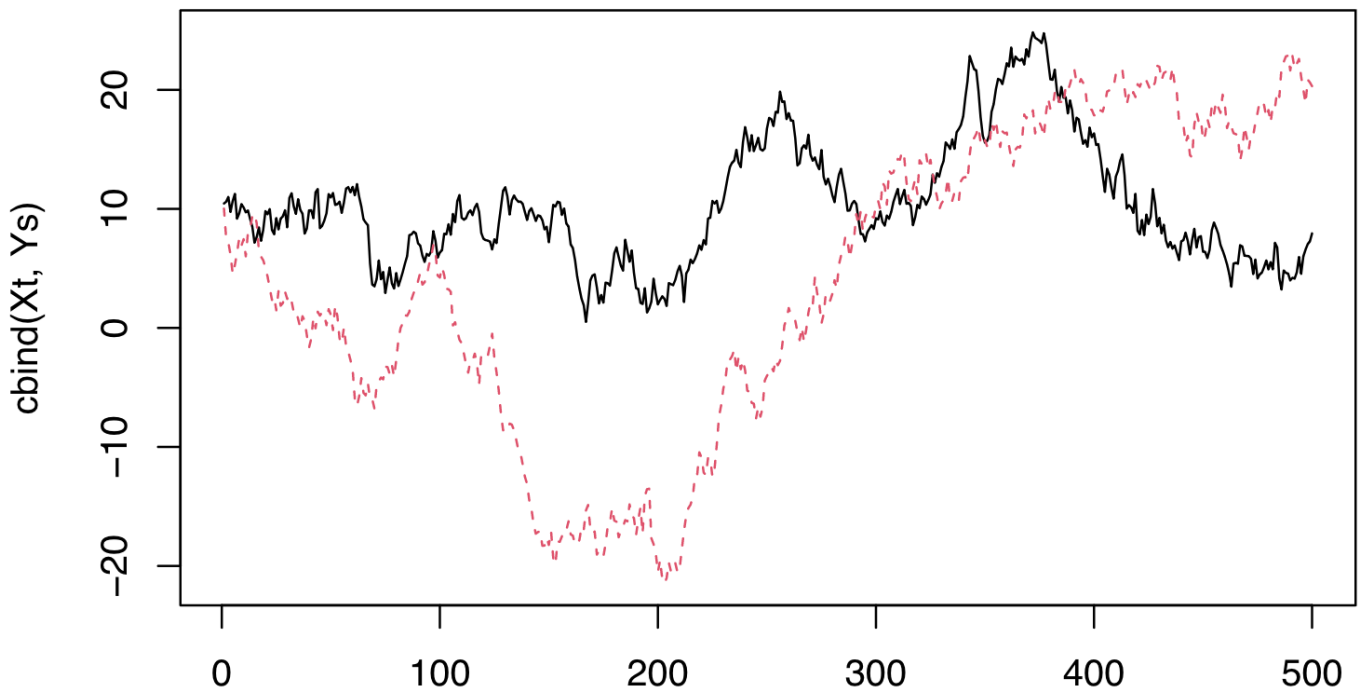
even though there should be no relationship between these two variables, the estimated model is likely to have:

- High R^2
- Significant value for β
- highly autocorrelated residuals
- Low DW statistic
- Very high variance for ϵ_t

This spurious relationship occurs because both variables are trended. The other features occur because, since both Y_t and X_t are $I(1)$, the error term ϵ_t is also $I(1)$. These results should not be taken seriously.

Spurious regression in R

Figure: Times series of two independent random walk processes



Spurious regression in R

```r
# spurious regression
spurMdl <- lm(Ys ~ 1 + Xt)

summary(spurMdl)
```

```
##
## Call:
## lm(formula = Ys ~ 1 + Xt)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -23.821 -10.062   0.311   8.887  24.404
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -5.4359     1.1913  -4.563 6.36e-06 ***
## Xt            0.9211     0.1032   8.924  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.99 on 498 degrees of freedom
## Multiple R-squared:  0.1379, Adjusted R-squared:  0.1361
## F-statistic: 79.64 on 1 and 498 DF,  p-value: < 2.2e-16
```

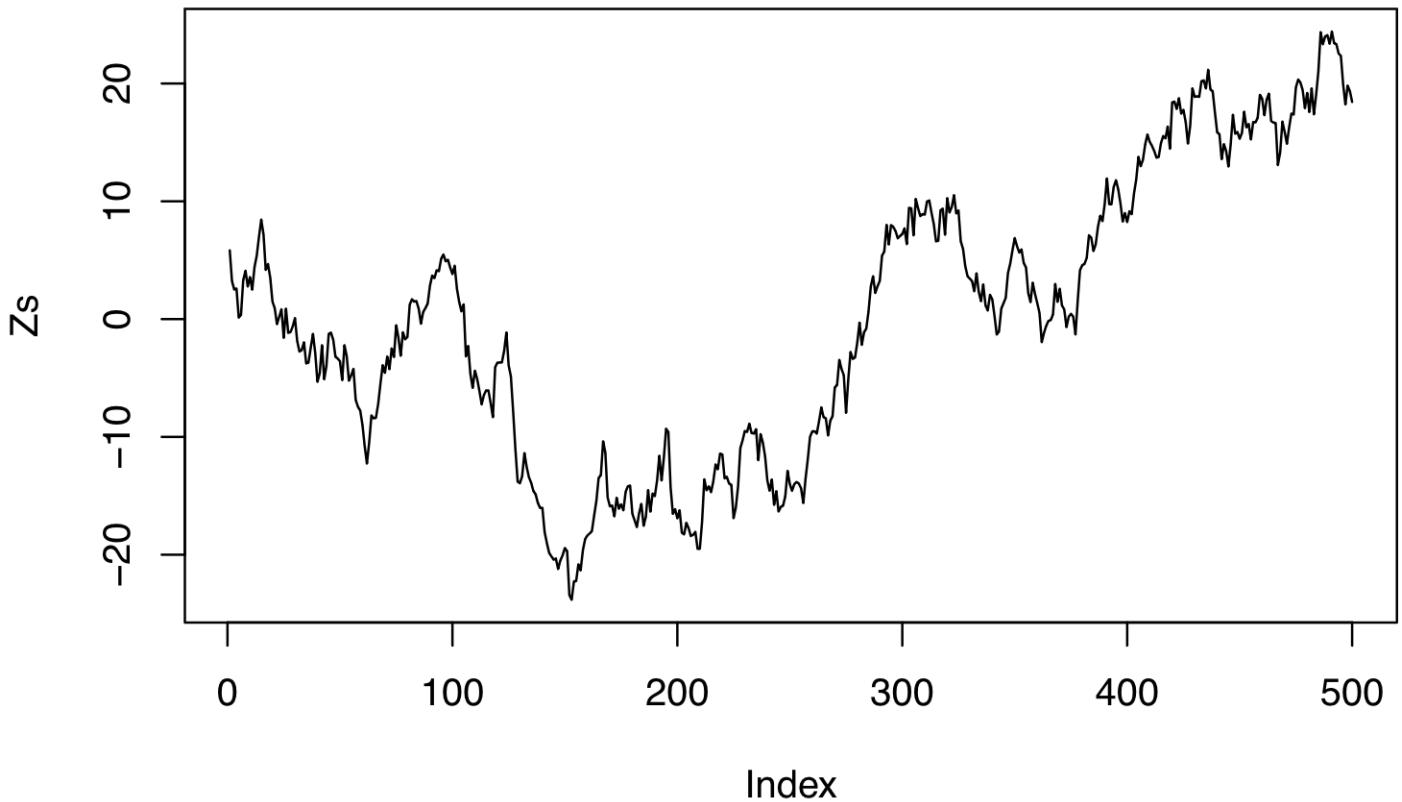
```r
# both constant and beta are highly significant

# DW Test
lmtest::dwtest(spurMdl)
```

```
##
##  Durbin-Watson test
##
## data:  spurMdl
## DW = 0.014127, p-value < 2.2e-16
## alternative hypothesis: true autocorrelation is greater than 0
```

Spurious regression in R

Figure: Times series Z_t from spurious regression




```
var(Zs) # very large variance as expected from non-sationary process
```

```
##          [,1]
## [1,] 143.4181
```

Cointegration

In the previous slide, we saw a case where Y_t and X_t were non-stationary and independent. Suppose instead that they share a common stochastic trend.

That is,

1. Y_t & X_t are $I(1)$
2. $\exists\ \beta$ such that $Y_t - \beta X_t$ is $I(0)$

In this case,

- we say Y_t and X_t are cointegrated and share a common trend
- $(1, -\beta)$ is called the cointegrating vector
- OLS estimate of β is said to be super consistent and hence, can be estimated by OLS.

Since $Z_t = Y_t - \beta X_t$ is $I(0)$ it must be that the long-run components of Y_t and βX_t cancel out over time (i.e., revert back to a mean of 0). This is related to the concept of a long-run equilibrium discussed previously.

Spurious regression in R

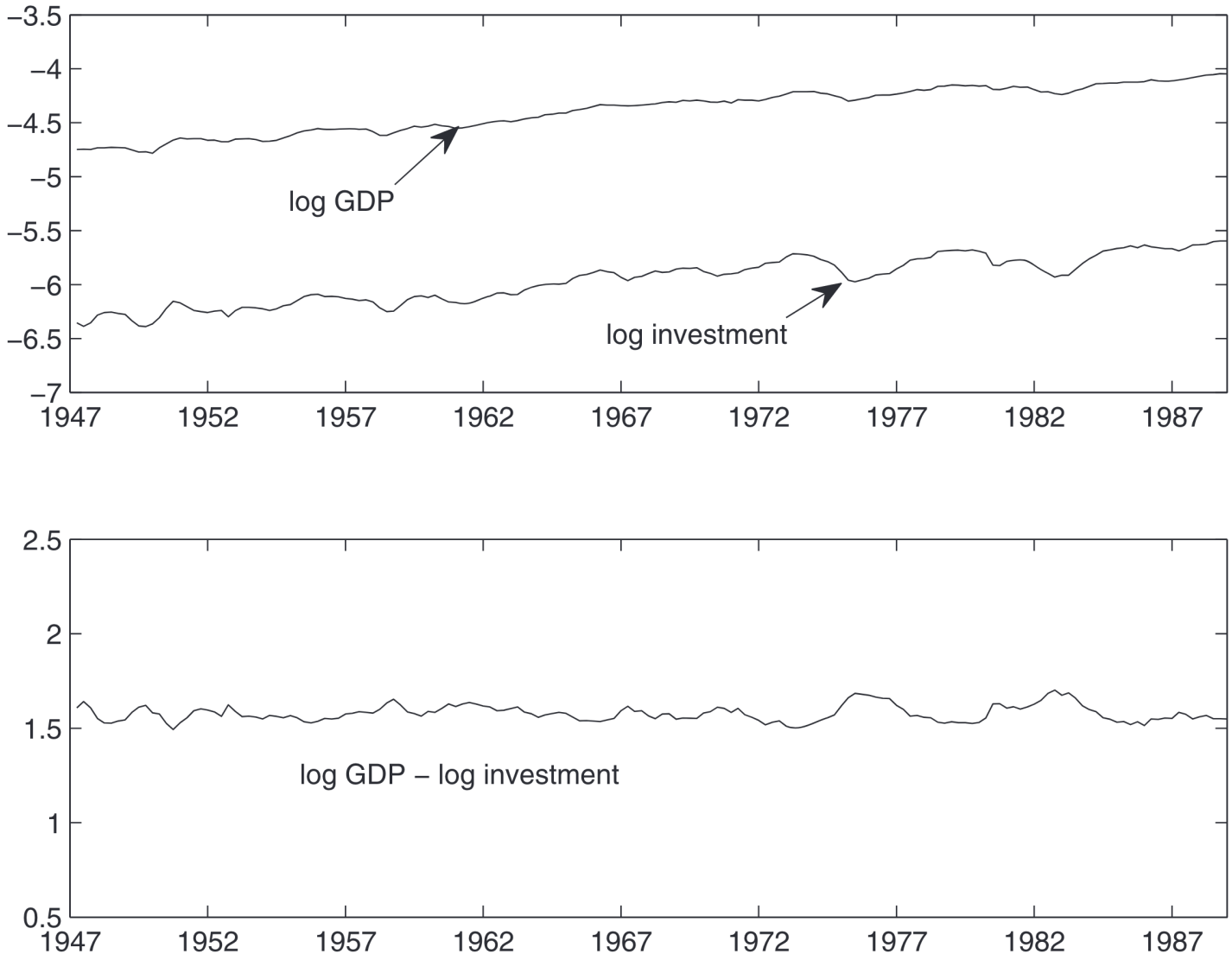


Figure 3.1. Logs of quarterly U.S. real GDP and real investment for 1947q1–1988q4.

From Kilian and Lütkepohl (2017), Ch. 3

Cointegration: Testing

So when Y_t and X_t are non-stationary and we estimate the regression

$$Y_t = \alpha + \beta X_t + \epsilon_t$$

Two outcomes are possible:

1. ϵ_t is $I(1)$
   - ▷ spurious regression
2. ϵ_t is $I(0)$
   - ▷ Y_t and X_t are cointegrated and share a common trend

Naturally, testing ϵ_t for a unit-root can help us determine which case we are dealing with. For this we can use

- Dickey-Fuller (DF) test for regression: $\Delta\hat{\epsilon}_t = \gamma_0 + \gamma_1\hat{\epsilon}_{t-1} + u_t$
- augmented DF (ADF) test by adding more lags of $\Delta\hat{\epsilon}_t$ to the above regression
- Cointegrating Regression Durbin-Watson (CRDW) test for dw statistic being close to 0

However, these test procedures have some limitations. See Verbeek (2004) for further discussion of these issues.

Cointegration: Error-correction Mechanism

Granger representation theorem, discussed in Granger (1983) and Engle and Granger (1987), states that if a set of variables are cointegrated, then a valid error-correction representation exists. The general form is given by

$$\theta(L)\Delta Y_t = \delta + \phi(L)\Delta X_{t-1} - \gamma Z_{t-1} + \alpha(L)\epsilon_t \tag{26}$$

where

- $Z_t = Y_t - \beta X_t$
- $(1, -\beta)$ is the cointegrating vector
- $\theta(L)$, $\phi(L)$, and $\alpha(L)$ are lag polynomials

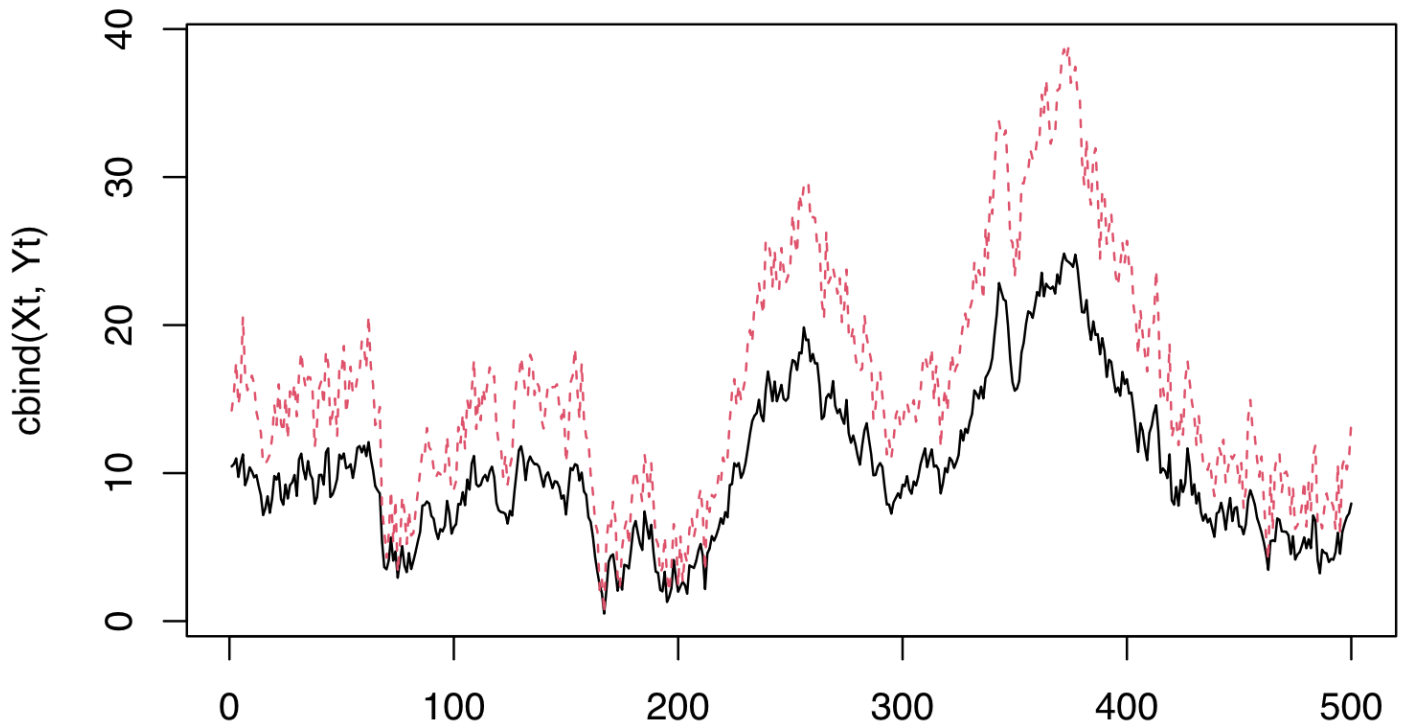
If β is known, we can compute Z_t directly and estimate this model. Otherwise, we use a two-step procedure where first we estimate $\hat{\beta}$ from the cointegrating regression

$$Y_t = \beta X_t + \epsilon_t \tag{27}$$

and then compute $\hat{Z}_t = Y_t - \hat{\beta} X_t$, which can be used in the error-correction representation regression

Cointegration regression in R

Figure: Times series of two cointegrated $I(1)$ processes



Cointegration regression in R

```r
# OLS estimation of beta
coinMdl <- lm(Yt ~ 1 + Xt)

summary(coinMdl)
```

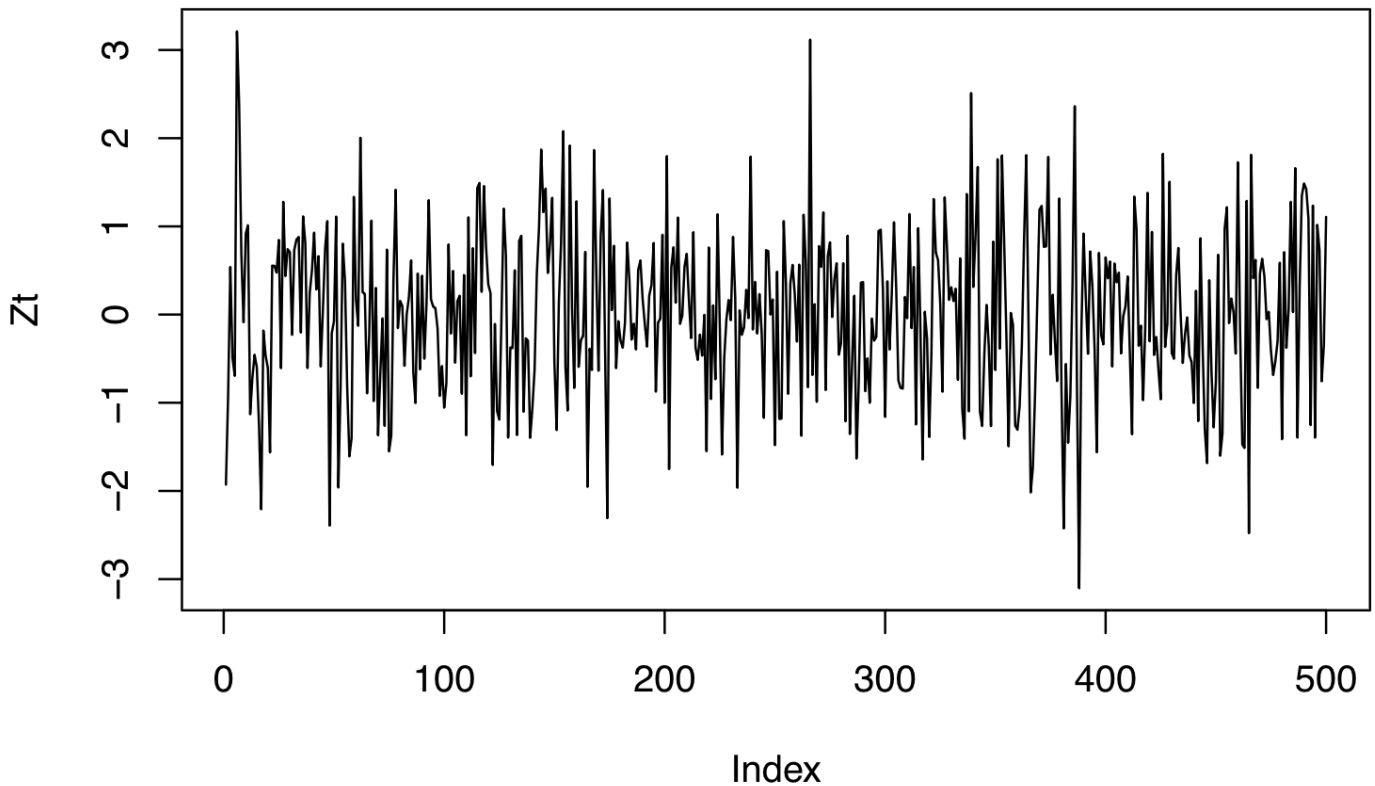
```
##
## Call:
## lm(formula = Yt ~ 1 + Xt)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.1022 -0.6064 -0.0069  0.6768  3.2092
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.371192   0.093749   3.959 8.61e-05 ***
## Xt          1.509978   0.008123 185.890  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9434 on 498 degrees of freedom
## Multiple R-squared:  0.9858, Adjusted R-squared:  0.9858
## F-statistic: 3.455e+04 on 1 and 498 DF,  p-value: < 2.2e-16
```

```r
# DW Test
lmtest::dwtest(coinMdl)
```

```
##
##  Durbin-Watson test
##
## data:  coinMdl
## DW = 1.9835, p-value = 0.4096
## alternative hypothesis: true autocorrelation is greater than 0
```

Cointegration regression in R

Figure: Times series Z_t from Cointegrated regression




```
var(Zt) # low variance close to 1 as expected
```

```
##             [,1]
## [1,] 0.8881854
```

Vector Autoregressive Models

Based on: Hamilton (1994) Ch. 10 & 11; Lütkepohl (2005) Ch. 2 - 4 & 6 - 8; Kilian and Lütkepohl (2017) Ch. 2 & 3; Candian (2021b)

See also: Verbeek (2004) Ch. 9; Enders (2015) Ch. 5 & 6; Martin et al. (2013) Ch. 13; Hayashi (2000) Ch. 6

Introduction

- Vector Autoregression (VARs) models were first introduced within the economics literature in Sims (1980) as a method for estimating large-scale macro models
- VAR models can be useful for:
  - Summarizing joint dynamics of economic time series data
  - Forecasting
  - Estimating causal relationships
  - Variance decomposition
  - Obtain Impulse Response Functions & Policy analysis
  - Test economic theories
- Two types of VAR models
  - Reduced-form VAR models
  - Structural VAR (SVAR) models (next section)

VAR(1) Example

We can model the joint dynamics of Federal Funds Rate (FFR) and GDP growth (ΔGDP) using the following VAR model:

$$\begin{aligned} \text{FFR}_t &= \phi_{11}\text{FFR}_{t-1} + \phi_{12}\Delta\text{GDP}_{t-1} + \epsilon_{f,t} \\ \Delta\text{GDP}_t &= \phi_{21}\text{FFR}_{t-1} + \phi_{22}\Delta\text{GDP}_{t-1} + \epsilon_{g,t} \end{aligned} \tag{28}$$

Here, the number of variables is $N = 2$, and the lag-order is $p = 1$.

When N and p are larger it is more convenient to work with matrix notation:

$$y_t = \Phi y_{t-1} + \epsilon_t \tag{29}$$

where

$$y_t = \begin{bmatrix} \text{FFR}_t \\ \Delta\text{GDP}_t \end{bmatrix} \qquad \Phi = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \qquad \epsilon_t = \begin{bmatrix} \epsilon_{f,t} \\ \epsilon_{g,t} \end{bmatrix}$$

VAR(p)

In general

$$y_t = \begin{bmatrix} y_{1,t} \\ y_{2,t} \\ \vdots \\ y_{N,t} \end{bmatrix}$$

and a VAR(p) model would be given by

$$y_t = \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \cdots + \Phi_p y_{t-p} + \epsilon_t \tag{30}$$

Note that here, the process y_t has been demeaned for ease of notation. That is, $y_t = Y_t - \boldsymbol{\mu}$, where $\boldsymbol{\mu}$ is an $(N \times 1)$ vector containing the mean of each variable in y_t.

Error Term

If we assume i.i.d. normal errors, then we can denote their joint distribution as

$$\epsilon_t \sim \mathcal{N}(\mathbf{0}, \Sigma_\epsilon) \qquad \text{where} \qquad \underset{(N\times N)}{\Sigma_\epsilon} = \begin{bmatrix} \sigma_{1,1}^2 & \sigma_{1,2} & \cdots & \sigma_{1,N} \\ \sigma_{2,1} & \sigma_{2,2}^2 & \cdots & \sigma_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{N,1} & \sigma_{N,2} & \cdots & \sigma_{N,N}^2 \end{bmatrix}$$

In the Federal Funds Rate and GDP growth example above where $N = 2$, the covariance matrix Σ_ϵ would be a simple (2×2) matrix given below

$$\Sigma_\epsilon = \begin{bmatrix} \sigma_f^2 & \sigma_{f,g} \\ \sigma_{g,f} & \sigma_g^2 \end{bmatrix}$$

Here $\sigma_{f,g} = \sigma_{g,f}$ is the covariance of FFR_t and ΔGDP_t

Companion Form

As can be seen from eq. (30), the VAR(p) model is simply a multivariate extension of the AR(p) model. Like AR(p) models, we can also use the companion form to obtain a VAR(1) representation:

$$x_t = Fx_{t-1} + \eta_t \tag{31}$$

where

$$\underset{(Np\times 1)}{x_t} = \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix} \quad \underset{(Np\times Np)}{F} = \begin{bmatrix} \Phi_1 & \Phi_2 & \dots & \Phi_{p-1} & \Phi_p \\ I & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I & 0 \end{bmatrix} \quad \underset{(Np\times 1)}{\eta_t} = \begin{bmatrix} \epsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

where I is $N \times N$ identity matrix.

It is often easier to work with the VAR(1) representation of the VAR(p) model obtained using the companion form. We can always recover y_t using $y_t = M_n x_t$ where M_n is an $N \times Np$ matrix defined as $M_n = [I, 0]$

Theoretical properties: Stationarity

The process y_t is second-order stationary iff the following conditions hold:

1. $\mathrm{E}[y_t^2] < \infty \; \forall t \in T$
2. $\mathrm{E}[y_s] = \mathrm{E}[y_t] \; \forall s, t \in T$
3. $\mathrm{Cov}(y_t, y_{t-h}) = \Gamma_{yy,h} \; \forall t \; \& \; \forall h$

If the process is second-order stationary we can also say it is covariance stationary or weakly stationary.

As in the univariate case, to determine if the process is stationary we can verify that the eigenvalues of F are inside the unit circle (i.e., less than 1 in absolute value or modulus) or equivalently if the roots of the characteristic equation are outside the unit circle

$$det\left[I - \Phi_1 z - \Phi_2 z^2 - \cdots - \Phi_p z^p\right] = 0 \tag{32}$$

For a proof of this see section 10.A of Hamilton (1994).

Theoretical properties: The mean

Suppose instead that we are working with the VAR(1) representation of a VAR(p) model that includes the constant term

$$x_t = \nu + Fx_{t-1} + \eta_t \tag{33}$$

where $\nu = [c, 0, \ldots, 0]'$ and c is the vector of constant term from the original VAR(p) model.

If the process is stationary, then we know $\mathrm{E}[y_t] = \mu \ \forall t$ and that $\mu = M_n \mu_x$ where $\mu_x = E[x_t]$. Using this, we can determine the mean as follows:

$$\begin{aligned} E[x_t] &= \nu + FE[x_{t-1}] + E[\eta_t] \\ \mu_x &= \nu + F\mu_x \\ (I - F)\mu_x &= \nu \\ \mu_x &= (I - F)^{-1}\nu \end{aligned}$$

and so $\mu = M_n(I - F)^{-1}\nu$

Theoretical properties: Autocovariances

Returning to the demeaned process for convenience and assuming stationarity holds, we can compute the autocovariance matrix of order zero as follows:

$$\begin{aligned}\Gamma_{xx,0} &= E[x_t x_t'] \\ &= F\Gamma_{xx,0}F' + E[\eta_t \eta_t']\end{aligned}$$

Using the $vec(\cdot)$ operator and $vec(ABC) = (C' \otimes A)vec(B)$ we can see that

$$\begin{aligned}vec(\Gamma_{xx,0}) &= vec(F\Gamma_{xx,0}F') + vec(E[\eta_t \eta_t']) \\ &= (F \otimes F)vec(\Gamma_{xx,0}) + vec(E[\eta_t \eta_t']) \\ &= [I - (F \otimes F)]^{-1} vec(E[\eta_t \eta_t'])\end{aligned}$$

by undoing the $vec(\cdot)$ (i.e., using $vec^{-1}(\cdot)$ with correct dimensions) we can obtain $\Gamma_{xx,0}$.

Theoretical properties: Autocovariances

Since $y_t = M_n x_t$ and $\Gamma_{xx,0} = E[x_t x_t']$, we can obtain $\Gamma_{yy,0}$ as follows:

$$\begin{aligned}\Gamma_{yy,0} &= E[y_t y_t'] \\ &= E[M_n x_t x_t' M_n'] \\ &= M_n \Gamma_{xx,0} M_n'\end{aligned}$$

Also, since

$$E[x_t x_{t-h}'] = FE[x_{t-1} x_{t-h}'] + E[\eta_t x_{t-h}']$$

we can see that (try to show this)

$$\Gamma_{xx,h} = F^h \Gamma_{xx,0} \tag{34}$$

and as before, $\Gamma_{y,h} = M_n \Gamma_{xx,h} M_n'$. Note also that $\Gamma_{yy,h} = \Gamma_{yy,-h}$.

Theoretical properties: Autocorrelations

Autocorrelation are useful because they are scale invariant.

The autocorrelation matrix for y_t is given by

$$R_{yy,h} = D^{-1}\Gamma_{yy,h}D^{-1} \tag{35}$$

where

$$D^{-1} = \begin{bmatrix} \frac{1}{\sqrt{\gamma_{11,0}}} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \frac{1}{\sqrt{\gamma_{NN,0}}} \end{bmatrix}$$

the autocorrelation of $y_{i,t}$ and $y_{j,t-h}$ is given by

$$\rho_{ij,h} = \frac{\gamma_{ij,h}}{\sqrt{\gamma_{ii,0}}\sqrt{\gamma_{jj,0}}} \tag{36}$$

which is simply the ij-element of the $R_{yy,h}$ matrix.

Theoretical properties: VMA(∞) representation

When stationary, we can represent a VAR(p) as a VMA(∞) as follows:

$$\begin{aligned} y_t &= \Phi_1 y_{t-1} + \cdots + \Phi_p y_{t-p} + \epsilon_t \\ (I - \Phi_1 L - \cdots - \Phi_p L^p)\, y_t &= \epsilon_t \\ \Phi(L) y_t &= \epsilon_t \end{aligned}$$

where $\Phi(L)$ is a finite order lag polynomial. As seen in Ch.8 of Verbeek (2004), the inverse of a finite order lag polynomial is an infinite order lag polynomial. Let $\Psi(L) = \Phi(L)^{-1}$. Then

$$\begin{aligned} \Psi(L)\Phi(L) y_t &= \Psi(L)\epsilon_t \\ y_t &= \Psi(L)\epsilon_t \\ y_t &= \sum_{j=0}^{\infty} \Psi_j \epsilon_{t-j} \end{aligned}$$

where $\sum_{j=0}^{\infty} |\Psi_j| < \infty$ and we define $\Psi_0 = I$. The result is also considered a causal VMA(∞) process.

Theoretical properties: VMA(∞) representation

We can also obtain the Covariance matrix using the VMA(∞) representation as

$$\begin{aligned}\Gamma_{yy,0} &= E[y_t y_t'] \\ &= E\left[\left(\sum_{j=0}^{\infty}\Psi_j\epsilon_{t-j}\right)\left(\sum_{j=0}^{\infty}\Psi_j\epsilon_{t-j}\right)'\right] \\ &= \sum_{j=0}^{\infty}\Psi_j\Sigma_\epsilon(\Psi_j)'\end{aligned}$$

Note that if the model include a constant term c then $\Psi(L)c = \mu$ and the last expression above becomes

$$y_t = \mu + \sum_{j=0}^{\infty}\Psi_j\epsilon_{t-j} \tag{37}$$

Estimation: OLS

Consider again the VAR(p) model given by eq. (30). And define the ($Np \times 1$) matrix $x_t = [y'_{t-1}, y'_{t-2}, \ldots, y'_{t-p}]'$. Now let,

$$Y = \begin{bmatrix} y'_1 \\ \vdots \\ y'_T \end{bmatrix} \quad X = \begin{bmatrix} x'_1 \\ \vdots \\ x'_T \end{bmatrix} \quad \underset{(Np \times N)}{\Phi} = \left[\Phi_1, \ldots, \Phi_p\right]' \quad \epsilon = \begin{bmatrix} \epsilon'_1 \\ \vdots \\ \epsilon'_T \end{bmatrix}$$

Using this notation, we can now write the VAR(p) model as

$$Y = X\Phi + \epsilon \tag{38}$$

From here, the consistent OLS estimator of Φ is given by

$$\hat{\Phi} = (X'X)^{-1}X'Y \tag{39}$$

and the sum of squared OLS residual matrix is given by

$$\hat{S} = (Y - X\hat{\Phi})'(Y - X\hat{\Phi}) \tag{40}$$

and a consistent estimator of Σ_ϵ is then given by

$$\hat{\Sigma}_\epsilon = \frac{1}{T}\hat{S} \tag{41}$$

See section 3.2.2 of Lütkepohl (2005) for asymptotic properties of the LS estimator.

Estimation: MLE

Since we assumed normality of the errors, we use the density of the multivariate normal distribution to estimate the model by Maximum Likelihood (ML) and so

$$f(y_t|x_t;\Phi,\Sigma_\epsilon) = (2\pi)^{\frac{-n}{2}}|\Sigma_\epsilon|^{\frac{-1}{2}}\exp\left\{\frac{-1}{2}(y_t-\Phi' x_t)'\Sigma_\epsilon^{-1}(y_t-\Phi' x_t)\right\}$$

by taking logs, we get

$$\log f(y_t|x_t;\Phi,\Sigma_\epsilon) = -\frac{n}{2}log(2\pi) - \frac{1}{2}log\left(|\Sigma_\epsilon|\right) - \frac{1}{2}(y_t-\Phi' x_t)'\Sigma_\epsilon^{-1}(y_t-\Phi' x_t)$$

this is the conditional log-likelihood function for observation y_t. It is conditional because are conditioning on all observations up to time t. When initializing the estimation in practice, we typically condition on the first p observations.

Estimation: MLE

For a sample $\{y_1, \ldots, y_T\}$ the conditional log-likelihood function is

$$\begin{aligned}\mathcal{L}(\theta) &= \sum_{t=1}^{T} \log f(y_t | x_t; \Phi, \Sigma_\epsilon) \\ &= -\frac{Tn}{2} log(2\pi) - \frac{T}{2} log\left(|\Sigma_\epsilon|\right) - \frac{1}{2} \sum_{t=1}^{T} \left(y_t - \Phi' x_t\right)' \Sigma_\epsilon^{-1} \left(y_t - \Phi' x_t\right)\end{aligned}$$

where $\theta = \{\Phi, \Sigma_\epsilon\}$

In this case, the MLE of Φ and Σ_ϵ and OLS estimators are equivalent (try to show this).

Lag-Order Selection

In the previous slides, we assumed the lag-order of the VAR model is known. I practice, we must choose p based on the available sample of data.

There are three main procedures for doing so:

- Top-down sequential testing
- Bottom-up sequential testing
- Information Criterion

Lag-Order Selection

- Top-down sequential testing
  1. Choose a maximum lag-order p_{max}
  2. Test hypothesis $H_0 : \Phi_{p_{max}} = 0$ vs. $H_a : \Phi_{p_{max}} \neq 0$
     - For example, this can be done using ML based tests such as Wald test or Likelihood ratio test
  3. If we fail to reject the null hypothesis then move on to testing $H_0 : \Phi_{p_{max}-1} = 0$ vs. $H_a : \Phi_{p_{max}-1} \neq 0$
  4. Continue sequential testing (removing one lag) until null hypothesis is rejected
  5. set the lag-order equal to the last alternative hypothesis from test where rejection occured
- Bottom-up sequential testing
- Information Criterion

Lag-Order Selection

- Top-down sequential testing
- Bottom-up sequential testing
  1. Choose a minimum lag-order p_{min}
  2. Estimate a VAR(p_{min}) model and perform residual autocorrelation test to see if model adequately captures dynamics in the data
     - ▷ Portmanteau Test for residual autocorrelation
     - ▷ LM Test for residual autocorrelation
  3. Continue to add more lags until residual autocorrelation test suggest model is adequate.
- Information Criterion

Lag-Order Selection

- Top-down sequential testing
- Bottom-up sequential testing
- Information Criterion
  - General form for information criterion:
    $C(p) = log\left(det\left(\hat{\Sigma}_\epsilon(p)\right)\right) + c_T\psi(p)$
    - $\hat{\Sigma}_\epsilon(p)$: estimate of the covar matrix when the lag-order is p
    - $\psi(p)$: function of order p which penalizes large lag orders
    - c_T: sequence of weights that may depend on sample size
  - Examples include"
    - Akaike Information Criterion (AIC):
      $AIC(p) = log\left(det\left(\hat{\Sigma}_\epsilon(p)\right)\right) + \frac{2}{T}(pN^2)$
    - Schwarz (Bayesian) Information Criterion (SIC) or (BIC):
      $BIC(p) = log\left(det\left(\hat{\Sigma}_\epsilon(p)\right)\right) + \frac{log(T)}{T}(pN^2)$
  - pN^2 is number of parameters. $pN^2 + N$ if constant term is included
  - Choose p that minimizes IC
  - Minimizing IC provides a balance between model fit and parsimony

VAR - Cointegration

In a VAR(p) model, the cointegrating vectors become cointegrating spaces of which the dimension is not known *a priori*.

The Granger representation theorem previously discussed extends to the vector setting. Starting with the VAR(p) model,

$$Y_t = \delta + \Phi_1 Y_{t-1} + \cdots + \Phi_p Y_{t-p} + \epsilon_t \tag{42}$$

The theorem states that there exists a valid error-correction representation. It is given by

$$\Delta Y_t = \delta + \Gamma_1 \Delta Y_{t-1} + + \cdots + \Gamma_{p-1} \Delta Y_{t-p+1} + \Pi Y_{t-1} + \epsilon_t \tag{43}$$

Here, the matrix $\Pi = -\Phi(L)$ is the long-run matrix, which determines the dynamic properties of Y_t

VAR - Cointegration

Here, $Y_t \sim I(1)$ and so $\Delta Y_t \sim I(0)$. By assumption we also have $\epsilon_t \sim I(0)$ and so it must be the case that $\Pi Y_{t-1} \sim I(0)$. This presents the following three settings

1. $\Pi = 0$
   - In this case, (43) is a stationary VAR$(p-1)$ model for ΔY_t
2. $Y_t \sim I(0)$
   - In this case, $\Pi = -\Phi(L)$ is full-rank and invertible and so we can write $Y_t = \Phi(L)^{-1}(\delta + \epsilon_t)$
3. If $rank(\Pi) = r$, $0 < r < N$
   - In this case, Π is rank deficient and there are linear combinations of ΠY_t that are stationary

This last case is the one of interest. In this case, there can be $n - r$ common trends shared among the variables.

VECM

In this case, we can write

$$\Pi = \underset{(N\times r)}{\gamma}\ \underset{(r\times N)}{\beta'} \tag{44}$$

where both, have $rank(\gamma) = rank(\beta') = r$. When using this in the (43), we get the vector error-correction model (VECM)

$$\Delta Y_t = \delta + \Gamma_1 \Delta Y_{t-1} + + \cdots + \Gamma_{p-1} \Delta Y_{t-p+1} + \gamma\beta' Y_{t-1} + \epsilon_t \tag{45}$$

where $\beta' Y_{t-1}$ gives the r cointegrating relationships and γ contains weights of these relationships and determine how elements of ΔY_{t-1} adjust to the r equilibrium errors. Hence,

$$Z_{t-1} = \beta' Y_{t-1} \tag{46}$$

Note that $\gamma\beta' = \gamma P^{-1} P\beta'$ for any invertible $(r \times r)$ matrix P. For this reason we only determine the space spanned by the columns of β'.

VECM - Testing

We can test for cointegration as follows:

- Engle-Granger Approach: regression of $Y_{1,t}$ on $Y_{2,t}, \ldots, Y_{N,t}$ and test for unit root in residuals. If null hypothesis of unit root is rejected, then we proceed to use $\hat{Z}_{t-1}$ in the VECM model
  - ▷ Results are sensitive to LHS variable (e.g., suppose cointegration vector does not include $Y_{1,t}$ but includes all other variables)
  - ▷ This test may have low power since it does not use all available information (other lags and interactions)
  - ▷ When more than one cointegration vector exists, OLS estimates a linear combination. Null hypothesis is no cointegration so this is not problematic.
- Johansen Type Test: Sequential procedure that avoids issues discussed above (see Johansen (1988); Johansen and Juselius (1990); Johansen (1991)).

VECM - Testing

We can test for cointegration as follows:

- Engle-Granger Approach: regression of $Y_{1,t}$ on $Y_{2,t}, \ldots, Y_{N,t}$ and test for unit root in residuals. If null hypothesis of unit root is rejected, then we proceed to use $\hat{Z}_{t-1}$ in the VECM model
- Johansen Type Test: Sequential procedure that avoids issues discussed above (see Johansen (1988); Johansen and Juselius (1990); Johansen (1991)).
  - Trace test: test $H_0 : rank(\Pi) = r_0$ vs. $H_1 : rank(\Pi) > r_0$ using

$$\lambda_{\text{trace}}(r_0) = -T \sum_{j=r_0+1}^{N} \log(1 - \hat{\lambda}_j)$$

  - Eigenvalue test: test $H_0 : rank(\Pi) = r_0$ vs. $H_1 : rank(\Pi) = r_0 + 1$ using

$$\lambda_{\max}(r_0) = -T \log(1 - \hat{\lambda}_{r_0+1})$$

  where $\hat{\lambda}_j$ are estimated eigenvalues of Π in decreasing order.

Limiting distribution of last two test statistics under null are approximated by a multivariate extensions of the Dickey–Fuller distributions.

What next?

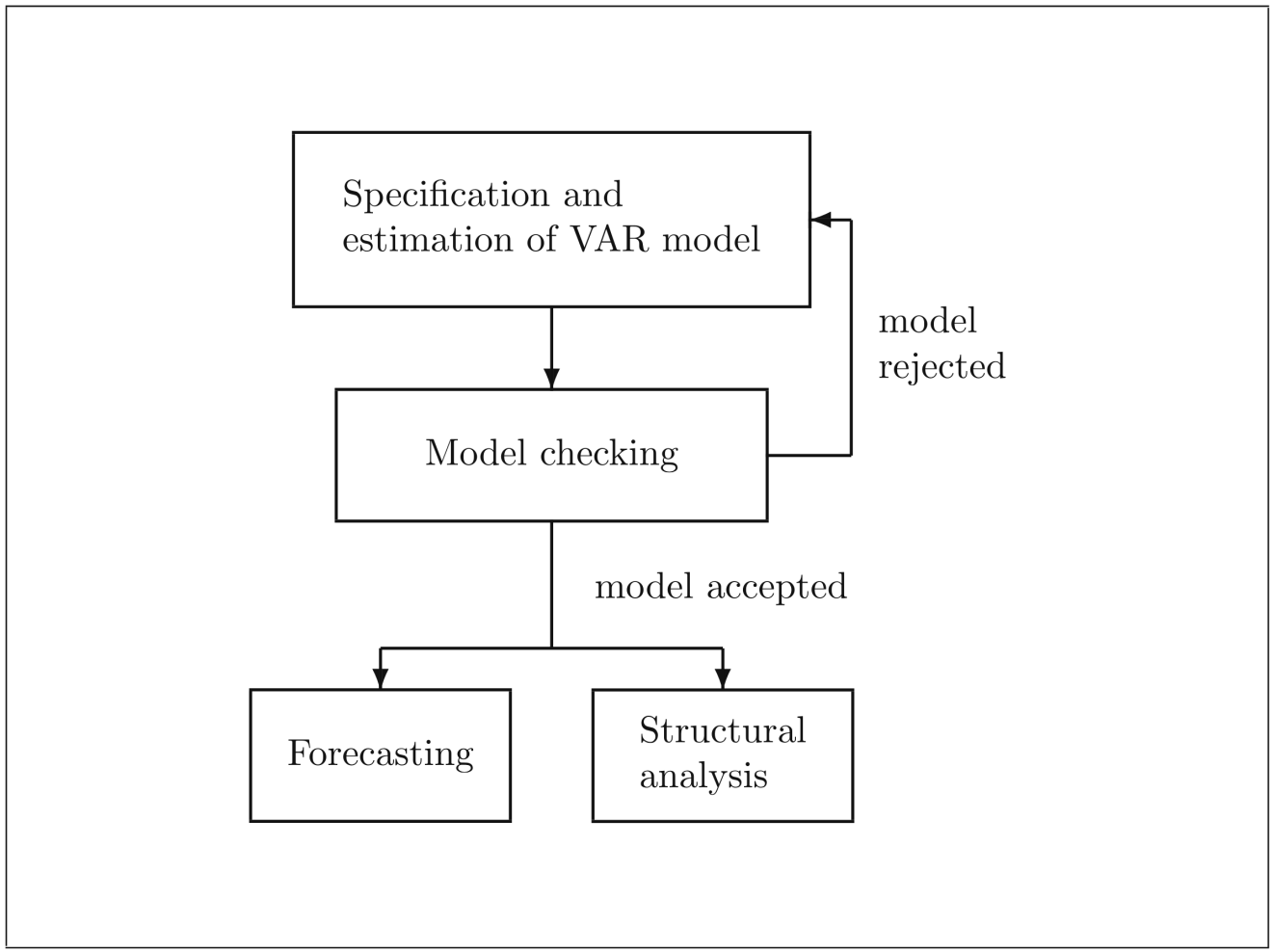


Fig. 1.1. VAR analysis.

From Lütkepohl (2005)

Prediction

Assuming we have an adequate N-dimensional stationary VAR model, we can use it to form some forecasts.

A good predictor will:

- Minimize a chosen Loss function
  - ▷ In practice we work with the expected loss
  - ▷ Typically the forecast Mean Square Error (MSE)
- Based on information available at time t (i.e., $\mathcal{I}_t$)

The linear minimum MSE predictor for horizon h at origin t is given by

$$E_t[y_{t+h}] \equiv E[y_{t+h}|\mathcal{I}_t] \tag{47}$$

where $\mathcal{I}_t = \{y_t, y_{t-1}, \dots\}$ is the information set containing all information up to time t. In this case, this linear MSE predictor is also an optimal predictor.

Prediction

Example: using VAR(1)

when $h = 1$

$$\begin{aligned} E_t[y_{t+1}] &= E_t[\Phi y_t] + E_t[\epsilon_{t+1}] \\ &= \Phi y_t \end{aligned}$$

when $h = 2$

$$\begin{aligned} E_t[y_{t+2}] &= \Phi E_t[y_{t+1}] + E_t[\epsilon_{t+2}] \\ &= \Phi^2 y_t \end{aligned}$$

when $h = 3$

$$\begin{aligned} E_t[y_{t+3}] &= E_t[\Phi y_{t+2}] + E_t[\epsilon_{t+3}] \\ &= \Phi^3 y_t \end{aligned}$$

In general

$$E_t[y_{t+h}] = \Phi^h y_t \tag{48}$$

Prediction

Let $\hat{y}_{t+h|t} = E_t[y_{t+h}]$. Then,

$$\begin{aligned} y_{t+h} - \hat{y}_{t+h|t} &= \Phi y_{t+h-1} + \epsilon_{t+h} - \Phi^h y_t \\ &= \Phi^2 y_{t+h-2} + \Phi \epsilon_{t+h-1} + \epsilon_{t+h} - \Phi^h y_t \\ &\vdots \\ &= \Phi^h y_t + \sum_{j=0}^{h-1} \Phi^j \epsilon_{t+h-j} - \Phi^h y_t \\ &= \sum_{j=0}^{h-1} \Phi^j \epsilon_{t+h-j} \end{aligned}$$

And so, the forecast error at horizon h is

$$\begin{aligned} \Sigma_y(h) &= E[(y_{t+h} - \hat{y}_{t+h|t})(y_{t+h} - \hat{y}_{t+h|t})'] \\ &= E\left[\left(\sum_{j=0}^{h-1} \Phi^j \epsilon_{t+h-j} \right) \left(\sum_{j=0}^{h-1} \Phi^j \epsilon_{t+h-j} \right)' \right] \end{aligned}$$

Prediction

And so

$$\Sigma_y(h) = \sum_{j=0}^{h-1} \Phi^j \Sigma_\epsilon (\Phi^j)' \tag{49}$$

$$= \sum_{j=0}^{h-1} \Psi_j \Sigma_\epsilon (\Psi_j)' \tag{50}$$

where the last line follows because, in the case of a VAR(1) model, $\Phi^j = \Psi_j$, the VMA coefficients obtained when inverting the lag polynomial $\Phi(L)$ and so $\Phi^0 = \Phi_0 = I$.

From here, we can see that when $h \to \infty$ we get

$$\Sigma_y(h) = \sum_{j=0}^{\infty} \Psi_j \Sigma_\epsilon (\Psi^j)'$$

$$= \Gamma_{yy,0}$$

That is, $\Sigma_y(h) \underset{h \to \infty}{\longrightarrow} \Gamma_{yy,0}$

Prediction

In general, if we have a VAR(p) model, we can work with its VAR(1) representation. In this case

$$\begin{aligned} y_{t+h} - \hat{y}_{t+h|t} &= M_n(x_{t+h} - \hat{x}_{t+h|t}) \\ &= M_n \left(\sum_{j=0}^{h-1} F^j \eta_{t+h-j} \right) \\ &= \sum_{j=0}^{h-1} M_n F^j M_n' M_n \eta_{t+h-j} \\ &= \sum_{j=0}^{h-1} \Psi_j \epsilon_{t+h-j} \end{aligned}$$

where $\hat{x}_{t+h|t} = F^h x_t$ was used.

Causality

Granger (1969) introduced a definition of causality that later became known as Granger causality.

Consider the following bi-variate VAR(1) model

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix} \tag{51}$$

Let $y_{1,t+h|t}$ denote the h-step ahead forecast of $y_{1,t}$ at origin t, based on information set $\mathcal{I}_t$ and the corresponding MSE for y_1 as $\sigma^2_{y_1}(h|\mathcal{I}_t)$.

We say $y_{2,t}$ Granger causes $y_{1,t}$ if

$$\sigma^2_{y_1}(h|\mathcal{I}_t) < \sigma^2_{y_1}(h|\mathcal{I}_t \backslash \{y_{2,s}|s \leq t\})$$

where $\mathcal{I}_t \backslash \{y_{2,s}|s \leq t\}$ represents the information set $\mathcal{I}_t$ except for the past and present information about $y_{2,t}$ process.

That is, if including past and present information about $y_{2,t}$ improves the forecast (reduces MSE) of $y_{1,t}$, then we say $y_{2,t}$ ranger causes $y_{1,t}$

Causality

In the bi-variate case, where $\mathcal{I}_t$ is limited to past and present observations of $y_{1,t}$ and $y_{2,t}$, it is straightforward to implement this concept using the VAR(p) model.

Using the model above, we can say $y_{2,t}$ does not Granger cause $y_{1,t}$ if and only if $\phi_{12} = 0$. The same can be said in a VAR(p) model if and only if $\phi_{12,j} = 0$ for $j = 1, 2, \ldots, p$.

These restrictions can easily be tested using a Wald or Likelihood Ratio test so long as the VAR(p) model is stationary.

Causality

For example, going back to the Federal Funds Rate and GDP growth model

$$\begin{bmatrix} \text{FFR}_t \\ \Delta\text{GDP}_t \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} \text{FFR}_{t-1} \\ \Delta\text{GDP}_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{f,t} \\ \epsilon_{g,t} \end{bmatrix} \tag{52}$$

we can test

$$H_0 : \phi_{21} = 0 \qquad \text{vs.} \qquad H_a : \phi_{21} \neq 0$$

If we fail to reject the null hypothesis, then we would say that our model suggests there is no evidence that Federal Funds Rate Granger causes GDP growth.

Causality

When we have a larger model (i.e., $N > 2$) it is no longer as straightforward to asses the Granger causality between two variables.

- Granger causal ordering may change when information set changes
- If $y_{2,t}$ Granger causes $y_{1,t}$, including other variables, say $y_{3,t}$, can break this causality if $y_{3,t}$ is correlated with both $y_{1,t}$ and $y_{2,t}$
- Similarly, $y_{2,t}$ may not Granger causes $y_{1,t}$ is a bi-variate model but does when including $y_{3,t}$.

For further discussion and extensions of Granger causality in higher dimensional VAR models see Lütkepohl (1993), Dufour and Renault (1998), Dufour et al. (2006), and Dufour and Taamouti (2010).

Impulse Response Functions (IRFs)

Continuing with the Federal Funds Rate and GDP growth model

$$\begin{bmatrix} \text{FFR}_t \\ \Delta\text{GDP}_t \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} \text{FFR}_{t-1} \\ \Delta\text{GDP}_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{f,t} \\ \epsilon_{g,t} \end{bmatrix} \tag{53}$$

Consider the following scenario

- $\text{FFR}_{t-1} = 0$ and $\Delta\text{GDP}_{t-1} = 0$
- At time t either $\epsilon_{f,t} = 1$ or $\epsilon_{g,t} = 1$
- Shocks are temporary so that after time t, $\epsilon_{f,s} = \epsilon_{g,s} = 0 \; \forall s > t$

IRFs will tell us how FFR_{t+h} and ΔGDP_{t+h} respond to that shock. In this case, we can define 4 IRFs. Namely,

$$\frac{\partial \text{FFR}_{t+h}}{\partial \epsilon_{f,t}} \qquad \frac{\partial \text{FFR}_{t+h}}{\partial \epsilon_{g,t}} \qquad \frac{\partial \Delta\text{GDP}_{t+h}}{\partial \epsilon_{f,t}} \qquad \frac{\partial \Delta\text{GDP}_{t+h}}{\partial \epsilon_{g,t}}$$

Impulse Response Functions (IRFs)

In general, using the VMA(∞) re presentation we derived earlier,

$$y_t = \sum_{j=0}^{\infty} \Psi_j \epsilon_{t-j}$$

moving forward h steps

$$y_{t+h} = \sum_{j=0}^{\infty} \Psi_j \epsilon_{t+h-j}$$

we can see that

$$\frac{\partial y_{t+h}}{\partial \epsilon_t} = \Psi_h$$

Let $e_1 = [1, 0, \ldots, 0]'$, the IRFs from a unit shock in $\epsilon_{1,t}$ is then

$$\frac{\partial y_{t+h}}{\partial \epsilon_{1,t}} = \Psi_h e_1$$

Impulse Response Functions (IRFs)

If we want to get the IRF of, say, the second variable, then we can define $e_2 = [0, 1, 0, \ldots, 0]'$ and we can obtain is as follows

$$\frac{\partial y_{2,t+h}}{\partial \epsilon_{1,t}} = e_2' \Psi_h e_1$$

Hence, the IRF of $y_{2,t+h}$ from a shock in $\epsilon_{1,t}$ is the $(2,1)$-element of Ψ_h and more generally, the the (i,j)-element of Ψ_h gives

$$\frac{\partial y_{i,t+h}}{\partial \epsilon_{j,t}} = \psi_{i,j}^{(h)} \qquad \text{where} \qquad \Psi_h = \begin{bmatrix} \psi_{1,1}^{(h)} & \cdots & \psi_{1,N}^{(h)} \\ \vdots & \ddots & \vdots \\ \psi_{N,1}^{(h)} & \cdots & \psi_{N,N}^{(h)} \end{bmatrix}$$

Note that, for example in a bivariate VAR(1) model, if $y_{j,t}$ does not Granger cause $y_{j,t}$, then $\phi_{i,j} = 0$ and so $\psi_{i,j}^{(h)} = 0$ and so the IRFs of i from shocks to j will be 0 $\forall h \geq 0$.

Impulse Response Functions (IRFs)

There is another method to get IRFs from a simple linear regression developed by Jordà (2005) called Local Projections. Consider the following regression

$$y_{t+h} = B_h y_t + \nu_{t+h}$$

It can be shown that $B_h = \Psi_h$ (Recall forecasting equation).

This method has gained popularity as it is believed to be robust to misspecification. There can be issues when process is highly persistent but this can be addressed with a lag-augmented model as discussed in Montiel Olea and Plagborg-Møller (2021).

Structural Vector Autoregressive Models

Based on: Lütkepohl (2005) Ch. 9 & 10; Kilian and Lütkepohl (2017) Ch. 4; Candian (2021a)

See also: Enders (2015) Ch. 5; Martin et al. (2013) Ch. 13; Hayashi (2000) Ch. 6

SVAR Model: Structural Shocks

So far we have been working with reduced-form VAR models

$$y_t = \Phi_1 y_{t-1} + \cdots + \Phi_p y_{t-p} + \epsilon_t$$

where $\epsilon_t \sim \mathcal{N}(0, \Sigma)$ and where ϵ_t has the interpretation of one-step ahead forecast errors and we allowed these errors to be correlated.

Correlated shocks make the IRFs difficult to interpret since a change in $\epsilon_{i,t}$ means $\epsilon_{j,t}$ also changes.

We can work with structural shocks (orthogonal) shocks instead. Let

$$\epsilon_t = A^{-1} u_t \tag{54}$$

So that the one-step-ahead forecast errors, ϵ_t, are linear combinations of the structural shocks u_t where $E[u_t u_t'] = I$ so that structural shocks are orthogonal.

SVAR Model: Contemporaneous Matrix

Using (54), we can express the VAR in structural form

$$y_t = \Phi_1 y_{t-1} + \cdots + \Phi_p y_{t-p} + A^{-1} u_t$$
$$A y_t = A\Phi_1 y_{t-1} + \cdots + A\Phi_p y_{t-p} + u_t$$

which gives the Structural VAR

$$A y_t = B_1 y_{t-1} + \cdots + B_p y_{t-p} + u_t \tag{55}$$

Here, the matrix A is the matrix of contemporaneous effects.

SVAR Models: IRFs

Using (54), we can also obtain the VMA(∞) in terms of the structural shocks

$$y_t = \sum_{j=0}^{\infty} \Psi_j A^{-1} u_{t-j} \tag{56}$$

and so the IRFs are now

$$\frac{\partial y_{t+j}}{\partial u_t} = \Psi_j A^{-1} \tag{57}$$

SVAR Models: Identification Problem

How do we identify A or A^{-1}?

Since $\epsilon_t = A^{-1}u_t$

$$\begin{aligned} E[\epsilon_t \epsilon_t'] &= E[A^{-1}u_t u_t'(A^{-1})'] \\ &= A^{-1}E[u_t u_t'](A^{-1})' \\ &= A^{-1}(A^{-1})' \\ &= \Sigma_\epsilon \end{aligned}$$

Consider for example a bivariate case and let $A^{-1} = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix}$. Then,

$$\begin{aligned} \sigma_1^2 &= \alpha_{11}^2 + \alpha_{12}^2 \\ \sigma_2^2 &= \alpha_{21}^2 + \alpha_{22}^2 \\ \sigma_{12} = \sigma_{21} &= \alpha_{11}\alpha_{21} + \alpha_{12}\alpha_{22} \end{aligned}$$

Since we know how to estimate Σ_ϵ, we might be tempted to solve for A^{-1}. However, we have 3 equations and 4 unknowns.

In general, A^{-1} has N^2 elements while Σ_ϵ only has $\frac{(N+1)N}{2}$ elements.

SVAR Models: Short-Run Restrictions

We know we must satisfy

$$\Sigma_\epsilon = E[A^{-1} u_t u_t' (A^{-1})']$$

one option for A^{-1} is the "square-root" (Cholesky decomposition) of Σ_ϵ.

$$\Sigma_\epsilon = \Sigma_{ltr}(\Sigma_{ltr})'$$

Σ_{ltr} is a lower triangular matrix and hence only has $\frac{(N+1)N}{2}$ elements. If Σ_ϵ is non-singular, this matrix is unique.

What does this mean for the dynamics of our model?

SVAR Models: Short-Run Restrictions

Working again with our example from Sims (1980)

$$\begin{bmatrix} \text{FFR}_t \\ \Delta\text{GDP}_t \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} \text{FFR}_{t-1} \\ \Delta\text{GDP}_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{f,t} \\ \epsilon_{g,t} \end{bmatrix} \tag{58}$$

Here, the shocks are monetary policy shocks ($\epsilon_{f,t}$) and technology shocks ($\epsilon_{g,t}$). We set $\alpha_{12} = 0$ so that $A^{-1} = chol(\Sigma_\epsilon)$. Then

$$\epsilon_t = \begin{bmatrix} \alpha_{11} & 0 \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \begin{bmatrix} u_{f,t} \\ u_{g,t} \end{bmatrix}$$

which lead to

$$\begin{aligned} \text{FFR}_t &= \phi_{11}\text{FFR}_{t-1} + \phi_{12}\Delta\text{GDP}_{t-1} + \alpha_{11}u_{f,t} \\ \Delta\text{GDP}_t &= \phi_{21}\text{FFR}_{t-1} + \phi_{22}\Delta\text{GDP}_{t-1} + \alpha_{21}u_{f,t} + \alpha_{22}u_{g,t} \end{aligned} \tag{59}$$

The assumption we have built into this model is: central bank does not react contemporaneously to technology shocks. This can be because data on aggregate output only become available with a one-quarter lag.

SVAR: Other Identification Methods

Depending on your application, short-run restrictions may not always be appropriate. Other important identifications methods include:

- Long-run restrictions (see Blanchard and Quah (1989))
- Sign restrictions (see Uhlig (2005))
- Heteroskedasticity (see Rigobon (2003))
  - ARCH/GARCH (see Normandin and Phaneuf (2004); Bouakez and Normandin (2010))
  - Markov-switching (see Lanne et al. (2010))
- Non-Gaussian errors (see Gouriéroux and Zakoïan (2015); Lanne et al. (2017)

References I

Blanchard, O. J. and Quah, D. (1989). The Dynamic Effects of Aggregate Demand and Supply Disturbances. *The American Economic Review*, 79(4):655 – 673.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31:307 – 327.

Bollerslev, T. (2009). Glossary to ARCH (GARCH). *School of Economics and Management- University of Aarhus. CREATES Research Paper 2008-49*.

Bouakez, H. and Normandin, M. (2010). Fluctuations in the foreign exchange market: How important are monetary policy shocks? *Journal of International Economics*, 81(1):139–153.

Candian, G. (2021a). Structural vector autoregressions. *ECON 8080207 Lecture notes, HEC Montreal*.

Candian, G. (2021b). Vector autoregressions. *ECON 8080207 Lecture notes, HEC Montreal*.

Diebold, F. X. and Mariano, R. S. (2002). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*, 20(1):134 – 144.

References II

Dufour, J.-M., Pelletier, D., and Renault, E. (2006). Short run and long run causality in time series: inference. *Journal of Econometrics*, 132(2):337–362.

Dufour, J.-M. and Renault, E. (1998). Short Run and Long Run Causality in Time Series: Theory. *Econometrica*, 66(5):1099.

Dufour, J.-M. and Taamouti, A. (2010). Short and long run causality measures: Theory and inference. *Journal of Econometrics*, 154(1):42–58.

Enders, W. (2015). *Applied econometric time series*. Wiley, Hoboken, NJ, fourth edition edition.

Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50(4):987 – 1007.

Engle, R. F. and Bollerslev, T. (1986). Modelling the persistence of conditional variances. *Econometric Reviews*, 5(1):1–50.

References III

Engle, R. F. and Granger, C. W. J. (1987). Co-Integration and Error Correction: Representation, Estimation, and Testing. *Econometrica*, 55(2):251 – 276.

Gouriéroux, C. and Zakoïan, J. (2015). On Uniqueness of Moving Average Representations of Heavy-tailed Stationary Processes. *Journal of Time Series Analysis*, 36:876–887.

Granger, C. W. (1983). *Co-integrated variables and error-correcting models*. PhD thesis, Discussion Paper 83-13. Department of Economics, University of California at ....

Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, Vol. 37(No. 3):424 – 438.

Gray, S. F. (1996). Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics*, 42(1):27–62.

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

References IV

Hamilton, J. D. and Susmel, R. (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of econometrics*, 64(1-2):307–333.

Hansen, P. R., Lunde, A., and Nason, J. M. (2011). The Model Confidence Set. *Econometrica*, 79(2):453–497.

Hayashi, F. (2000). *Econometrics*. Princeton University Press.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 12(2-3):231–254.

Johansen, S. (1991). Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models. *Econometrica*, 59(6):1551.

Johansen, S. and Juselius, K. (1990). Maximum likelihood Estimation and inference on cointegration — WITH applications to the demand for money. *Oxford Bulletin of Economics and Statistics*, 52(2):169–210.

Jordà, O. (2005). Estimation and inference of impulse responses by local projections. *American Economic Review*, 95(1):161–182.

References V

Kilian, L. and Lütkepohl, H. (2017). *Structural Vector Autoregressive Analysis:*. Cambridge University Press, 1 edition.

Lanne, M., Lütkepohl, H., and Maciejowska, K. (2010). Structural vector autoregressions with Markov switching. *Journal of Economic Dynamics and Control*, 34(2):121–131.

Lanne, M., Meitz, M., and Saikkonen, P. (2017). Identification and estimation of non-Gaussian structural vector autoregressions. *Journal of Econometrics*, 196(2):288–304.

Lütkepohl, H. (1993). Testing for causation between two variables in higher dimensional var models. *Studies in Applied Econometrics, ed. by H. Schneeweiss and K. Zimmerman. Heidelber*.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. New York : Springer.

Martin, V., Hurn, S., and Harris, D. (2013). *Econometric Modeling with Time Series: Specification, Estimation and Testing*. Themes in Modern Econometrics. Cambridge University Press.

References VI

Montiel Olea, J. L. and Plagborg-Møller, M. (2021). Local Projection Inference Is Simpler and More Robust Than You Think. *Econometrica*, 89(4):1789–1823.

Nelson, D. B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica*, 59(2):347 – 370.

Normandin, M. and Phaneuf, L. (2004). Monetary policy shocks: Testing identification conditions under time-varying conditional volatility. *Journal of Monetary Economics*, 51(6):1217–1243.

Rigobon, R. (2003). Identification Through Heteroskedasticity. *Review of Economics and Statistics*, 85(4):777–792.

Sims, C. A. (1980). Macroeconomics and Reality. *Econometrica*, 48(1):1.

Uhlig, H. (2005). What are the effects of monetary policy on output? Results from an agnostic identification procedure. *Journal of Monetary Economics*, 52(2):381–419.

Verbeek, M. (2004). *A Guide to Modern Econometrics*. John Wiley & Sons, Ltd, 2nd edition.